

**УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ  
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б.Н.ЕЛЬЦИНА**

На правах рукописи

**ЧЕРНАВИН ФЕДОР ПАВЛОВИЧ**

**МЕТОДИЧЕСКИЙ ИНСТРУМЕНТАРИЙ  
ОЦЕНКИ И МОДЕЛИРОВАНИЯ КРЕДИТНОГО РИСКА  
ПО ПОТРЕБИТЕЛЬСКИМ КРЕДИТАМ  
С ПРИМЕНЕНИЕМ КОМИТЕТНЫХ КОНСТРУКЦИЙ**

Специальность 08.00.13 – «Математические и инструментальные методы  
экономики»

**ДИССЕРТАЦИЯ**

на соискание учёной степени  
кандидата экономических наук

Научные руководители:  
доктор физико-математических наук, профессор  
Никонов Олег Игоревич

доктор экономических наук, профессор  
Акбердина Виктория Викторовна

**ЕКАТЕРИНБУРГ – 2016**

## Содержание

Введение .....	3
Глава 1. Капитал кредитной организации и методы оценки кредитных рисков.....	11
1.1. Подходы к классификации рисков кредитной организации. Капитал как основной источник погашения риска.....	11
1.2. Методы оценки кредитного риска .....	27
Глава 2. Применение комитетных конструкций к классификации заемщиков.....	42
2.1. Классификация заемщиков физических лиц. Формулировка задачи классификации заемщиков методом комитета .....	42
2.2. Метод комитетов. Формулировка задачи принятия решения методом комитетов.....	44
2.3. Построение комитета через решение задачи математического программирования.....	47
2.3 Выбор целевой функции. Сравнение результатов комитетных решений в зависимости от выбранной целевой функции.....	64
2.4 Оценка сложности построения комитета. Нахождение начального решения задачи частично-целочисленного программирования .....	73
Глава 3. Построение рейтинговой модели оценки заемщика на основе комитетных конструкций .....	80
3.1. Данные и оценка параметров заемщика .....	80
3.2. Построение рейтинговой модели оценки заемщика.....	90
3.3. Валидация рейтинговых моделей. ....	116
3.4. Оценка экономического эффекта от внедрения рейтинговой модели.....	123
3.5. Алгоритмы и принципы применения рейтинговой модели ....	127
3.6. Применение комитетных конструкций к принятию решений на финансовых рынках.....	131
Заключение.....	139
Библиографический список.....	147

## **Введение**

**Актуальность темы исследования.** Российский банковский сектор в 2014-2015 годах столкнулся с рядом проблем, связанных с недооценкой кредитных рисков по потребительским кредитам, выданным в 2011-2013 годах. Так, за 2014 год остатки по резервам на возможные потери по кредитам физических лиц возросли с 619,0 млрд. руб. до 939,0 млрд. руб. (+ 320,0 млрд. руб., + 51,7%), уровень покрытия резервами увеличился с 6,5% до 8,6% (+ 2,1 п.п.). В настоящее время Банком России осуществляется внедрение подходов к оценке качества активов и капитала кредитной организации, определенных в рамках «Международной конвергенции измерения капитала и стандартов капитала: новые подходы» (Базель II). Основными нововведениями Базеля II оценке рисков, состоит в допуске внутренних рейтингов к расчету достаточности капитала кредитной организации (IRB approach). Кредитные организации заинтересованы как в снижении кредитных рисков, так и к разработке систем оценки рисков, которые бы удовлетворяли требованиям IRB подхода, что позволило бы снизить нагрузку на капитал кредитной организации. Одним из основных требований к подходу к оценке рисков на основании внутренних рейтингов, является внедрение в кредитной организации рейтинговой системы, которая бы позволяла с высокой степенью уверенности дифференцировать заемщиков по степени риска.

Чаще всего к построению рейтинговых моделей применяются методы построения линейного дискриминанта (как например, Logit метод), при этом данные методы, не позволяют учитывать нелинейные зависимости параметров, характеризующих заемщика. Актуальной проблемой является нахождение методов, позволяющих учитывать нелинейные зависимости кредитного риска и параметров заемщиков. Комитетные конструкции являются одним из методов, позволяющих проводить классификацию с учетом нелинейных зависимостей переменных, заметим, что данный метод не нашел широкого применения к построению рейтинговых моделей. Основным плюсом метода комитетов по

сравнению с иными методами классификации, такими как, например нейронные сети, является наличие содержательной и геометрической интерпретации решения, так метод комитетов можно представить, как модель кредитного комитета из  $N$  членов.

**Степень изученности проблемы.** В области регулирования кредитного риска системообразующими документами являются Международная конвергенция измерения капитала и стандартов капитала: новые подходы (Базель II), положением Банка России №254-п и письмом №192-Т. Оценка кредитных рисков в банке имеет следующие основные цели: оценка рисков для определения ставок по кредитным продуктам, в том числе дифференциация ставок в рамках кредитного продукта, определение требований к капиталу кредитной организации, снижение волатильности финансового результата кредитной организации. Передовая практика в решении данных задач рассматривается в информационных документах зарубежных регуляторов и Базельского комитета по банковскому надзору.

Вопросы применения современных подходов (продвинутых методов) к оценке кредитного риска и капитала кредитной организации освещены в работах следующих зарубежных авторов: А. Хамерле, Т. Либих, Д. Реш, Э. Алтман, А. Рести, А. Сирони, Е. Анжелини, Л. Аллен, А. Сандерс, Ж. Крука, Д. Эделман, Л. Томас, Ч. Гудхард, М. Сеговиано, М. Горди, Т. Джакобсон, Ж. Линде, К. Росбах. Анализ применения методов машинного обучения приведен в работах: Д. Харда, В. Хенлей, А. Атия, Р. Филдес, К. Николополос и других. Отмечается значительный рост работ по методам анализа риска со стороны китайских, тайваньских и южнокорейских авторов, таких как: Ю. Ким, С. Сон, Л. Юа, Ш. Ван, К. Либо, Т. Ли, Ч. Чиюб.

Значительный вклад в теорию оценки банковских рисков и взаимосвязанности рисков с пруденциальными подходами к оценке рисков внес А. Ю. Симановский. Тема о связи капитала банка и кредитного риска поднимается в работах Е. С. Карповой, Е. В. Травкиной, А. В. Урсуленко. Вопросы анализа рисков, в том числе банковских, рассматриваются в рамках

научной школы О. И. Никонова. Вопросы применения методов машинного обучения к оценке кредитного риска рассматриваются в работах В. Е. Селянина, А. М. Порошиной, Я. С. Мязовой, А. В. Моисеева, Е. А. Поправко, Н. Г. Федотова.

Понятие комитета было введено К. Аблоу и Д. Кейлором в 1965 году в работе «Inconsistent Homogenous Linear Inequalities». Дальнейшее развитие метод комитетов получил Екатеринбургской школе распознавания образов Института математики и механики в работах Вл. Д. Мазурова, М. Ю. Хачая, А. И. Кривоногова, Н. Г. Белецкого, А. И. Смирнова. В рамках данных работ была показана применимость комитетных конструкции к решению практических задач классификации и распознавания образов. Заметим, что тема применения комитетных конструкций к построению моделей оценки кредитного риска является не изученной как российскими, так и зарубежными авторами.

**Объект исследования** – финансовые организации, реализующие процедуры оценки кредитных рисков по потребительским кредитам.

**Предмет исследования** – экономические отношения, возникающие по поводу управления риском кредитования в коммерческом банке.

**Основная гипотеза.** К оценке кредитного риска можно применить один из методов классификации – метод комитетов, который ранее не использовался для оценки кредитных рисков. Предполагается, что существует зависимость между числом членов комитета, «голосующих» за возникновение дефолта по кредитному договору, и кредитным риском. На основе комитетных конструкций могут быть разработаны модели оценки кредитного риска, удовлетворяющие требованиям Базельских соглашений.

**Цель исследования** состоит в разработке методического инструментария моделирования кредитного риска по потребительским кредитам на основе комитетных конструкций.

Для достижения этой цели были поставлены **следующие задачи:**

1) Обосновать методический подход и разработать инструментарий построения комитетных конструкций через решение задачи частично целочисленного программирования для целей классификации заемщиков банка.

2) Разработать рейтинговую модель оценки вероятности дефолта заемщика на основе комитетных конструкций.

3) Разработать методику оценки экономического эффекта применения рейтинговой модели на основе комитетных конструкций, обосновать алгоритм последовательной оптимизации комитетного решения, что позволяет проводить корректировку весов рейтинговой модели.

**Теоретико-методологической базой** исследования послужили положения теории комитетных конструкций, теории математического программирования, теории вероятностей, теории распознавания образов и классификации. Методическую основу исследования составили нормативные документы Банка России и материалы Базель II.

**Достоверность и обоснованность подходов и выводов** подтверждается достаточным объемом статистических данных о заемщиках (общее число проанализированных кредитных договоров – свыше 60 тыс. единиц), использованием методов математического и имитационного моделирования, математической статистики.

**Информационную базу исследования** составили данные о кредитных договорах, предоставленные кредитной организацией, положения, письма и статистические данные Банка России, методики оценки кредитных рисков, используемые российскими и зарубежными банками, а также сведения, содержащиеся в публикациях отечественных и зарубежных авторов.

**Основные методы исследования.** В диссертационном исследовании использовались общенаучные методы анализа (сравнительного, структурного, системного, логического). Применены методы линейного программирования, математического моделирования, эвристические методы поиска решения (эволюционный алгоритм и последовательная оптимизация решений), также в работе использованы методы математической статистики и теории вероятностей,

использовались пакеты прикладных программ: IBM ILOG CPLEX, IBM SPSS Statistics, Wolfram Mathematica. Эволюционные алгоритмы реализованы на языке Visual Basic for Application.

**Основные научные результаты, полученные лично автором, и их научная новизна:**

1) Обоснован методический подход классификации заемщиков банка, базирующийся на теории комитетных конструкций, подразумевающий под членом комитета полупространство в пространстве социально-экономических признаков заемщика, совокупность которых позволяет принять решение о выдаче кредита. На основе методического подхода разработан авторский инструментарий, включающий комитеты большинства, единогласия, р-комитет и комитет с неравными весами членов, учитывающие процедуры голосования по совокупности признаков заемщика, и отличающийся классификацией заемщиков на основе решения задач линейного частично-целочисленного программирования, которые ранее не применялись для построения комитетных конструкций, что позволяет в отличие от традиционных Logit и Probit моделей учесть нелинейные зависимости переменных, в отличие от нейронных сетей – получить высокую интерпретируемость результатов решения (п.1.1. Паспорта специальности 08.00.13 ВАК РФ).

2) Обоснована возможность применения комитетных конструкций для построения рейтинговых моделей оценки заемщиков, в частности эмпирически подтверждена гипотеза о зависимости риска дефолта заемщика от числа членов комитета, «проголосовавших» за дефолт по кредитному договору. На основе доказанной гипотезы разработана авторская рейтинговая модель оценки вероятности дефолта заемщика, отличающаяся от традиционных рейтинговых моделей использованием комитетных конструкций на базе р-комитета и комитета с разными весами, включающая частные рейтинговые модели для множеств заемщиков с низким и высоким уровнями риска, предполагающая выделение рейтинговых групп на основе числа членов комитета, «голосовавших» за дефолт, и подразумевающая валидацию на основе скорректированного индекс

Херфиндаля, мер AUROC и AR. (п.1.2, 1.4 Паспорта специальности 08.00.13 ВАК РФ).

3) Разработана методика оценки экономического эффекта от классификации заемщиков с применением авторской рейтинговой модели, основанная на оценке высвобождаемого капитала банка, включающая расчет взвешенных по риску активов и требований к капиталу, отличающаяся от традиционных методик высокой чувствительностью к социально-экономическим признакам заемщика за счет применения комитетных конструкций, что позволяет количественно оценить кредитные риски банка, обосновать увеличение кредитного портфеля за счет снижения требований к капиталу и получение дополнительной прибыли. Авторская методика включает также алгоритм последовательной оптимизации комитетного решения, что позволяет проводить корректировку весов рейтинговой модели (п.1.6 Паспорта специальности 08.00.13 ВАК РФ). Теоретическая значимость исследования состоит в развитии теории комитетных конструкций в части применения ее для оценки и моделирования кредитных рисков. Практическая значимость результатов диссертации заключается в разработке прикладного инструментария для оценки и моделирования кредитного риска в целях улучшения качества активов коммерческого банка и снижения требования к капиталу.

#### **Апробация результатов исследования:**

Работа автора, связанная с применением метода комитетов к оценке кредитных рисков вышла в финал «Второго национального конкурса по экономике» и отмечена наградой министра экономического развития РФ.

Основные результаты диссертационной работы докладывались и получили положительную оценку на международных и всероссийских научных конференциях: Всероссийская научно-практическая конференция АИСТ'2013 «Анализ Изображений, Сетей и Текстов», Екатеринбург, 4–6 апреля 2013; X Международная научная конференция по проблемам экономического развития в современном мире «Устойчивое развитие российских регионов: Россия и ВТО», Екатеринбург, УРФУ, 19-20 апреля 2013; IV Международная научно-



практическая конференция «Социально-экономическое развитие регионов России», Москва, МЭСИ, 23 мая 2014; XIII Международная молодежная конференция «Новые тенденции в экономике и управлении организацией», Екатеринбург, УРФУ, 25 – 27 июня 2014; Вл.Д. Мазуров и Уральская научная школа распознавания образов, Екатеринбург, УИЭУиП, 12 марта 2015; Устойчивое развитие российских регионов: экономическая политика в условиях внешних и внутренних шоков, Екатеринбург, УРФУ, 17-18 апреля 2015.

Основные результаты работы использованы ПАО «УралТрансБанк» и внедрены в модель принятия решений по кредитным заявкам, использованы Уральским банком ПАО «Сбербанк» для формирования бизнес-плана на 2015-2016 года, внедрены в тестовом режиме в скоринговую модель оценки заемщика в ООО Банк «Нейва». Результаты внедрения подтверждены актами.

**Публикации.** Основные положения диссертации опубликованы в 13 публикациях общим объемом 6,56 пл. (авторских 4,1 п.л.), из них – 3 статьи в журналах, входящих в перечень изданий, рекомендованных ВАК («Деньги и Кредит» (учредитель журнала - Центральный банк Российской Федерации) – 2 статьи, «Экономика и предпринимательство» - 1 статья), 9 статьями в материалах конференций.

**Структура диссертации.** Диссертация состоит из введения, 3-х глав, заключения, списка литературы из 190 наименований. Основной объем работы составляет 166 страниц машинописного текста, включает 49 таблиц и 39 рисунков.

Во введении обоснована актуальность исследования, отражена степень изученности проблем, описаны цели и задачи диссертационной работы, перечислены основные результаты.

Первая глава «Капитал кредитной организации и методы оценки кредитных рисков» носит обзорный характер. В ней описаны основные подходы к оценке рисков кредитных организаций, пруденциальные подходы к оценке рисков и капитала. Рассмотрена классификация методов оценки риска, в рамках которой комитетные конструкции включены в группу классификационных методов

оценки риска. Проведено сравнение комитетных конструкций с иными классификационными методами оценки заемщика.

Во второй главе «Применение комитетных конструкций к классификации заемщиков» сформулирована задача классификации заемщиков методом комитетов и показано сведение задачи построения комитетов к задачам линейного частично-целочисленного программирования, определена целевая функция задачи. Проведен расчет сложности решения задачи и показана возможность применения эволюционного алгоритма для поиска начального решения задачи линейного частично-целочисленного программирования.

В третьей главе «Построение рейтинговой модели оценки заемщика на основе комитетных конструкций» дается описание параметров заемщиков и вводятся новые параметры, показывается возможность применения метода главных компонент для уменьшения числа параметров, описывающих заемщика, что снижает вычислительную сложность задачи. Рассматривается гипотеза о зависимости кредитного риска от числа членов комитета, проголосовавших за дефолт заемщика. Производится построение рейтинговых моделей и сведение рейтинговых моделей в итоговую модель оценки вероятности дефолта заемщика, оценивается качество рейтинговой модели валидационными тестами. Рассчитывается экономический эффект от внедрения моделей и разрабатывается методология дальнейшего применения рейтинговой модели.

# **Глава 1. Капитал кредитной организации и методы оценки кредитных рисков**

## **1.1. Подходы к классификации рисков кредитной организации. Капитал как основной источник погашения риска**

### *1.1.1 Классификация рисков кредитной организации*

Риск является неотъемлемой характеристикой банковской деятельности. Наиболее в широком смысле риском можно считать вероятность возникновения события, в результате которого возникает невозможность определения результатов деятельности кредитной организации (финансовых вложений, кредитной деятельности и других).

В письме Банка России от 23 июня 2004 г №70-Т "О типичных банковских рисках" [7] к типичным банковскими рисками относятся:

Кредитный риск - риск возникновения у кредитной организации убытков вследствие неисполнения, несвоевременного либо неполного исполнения должником финансовых обязательств перед кредитной организацией в соответствии с условиями договора. В работах [20,23,24,54,55,57,66,68] рассматриваются вопросы управления и анализа кредитного риска.

Страновой риск (включая риск неперевода средств) - риск возникновения у кредитной организации убытков в результате неисполнения иностранными контрагентами (юридическими, физическими лицами) обязательств из-за экономических, политических, социальных изменений, а также вследствие того, что валюта денежного обязательства может быть недоступна контрагенту из-за особенностей национального законодательства (независимо от финансового положения самого контрагента).

Рыночный риск - риск возникновения у кредитной организации убытков вследствие неблагоприятного изменения рыночной стоимости финансовых инструментов торгового портфеля и производных финансовых инструментов

кредитной организации, а также курсов иностранных валют и (или) драгоценных металлов.

Рыночный риск включает в себя фондовый риск, валютный и процентный риски.

Фондовый риск - риск убытков вследствие неблагоприятного изменения рыночных цен на фондовые ценности (ценные бумаги, в том числе закрепляющие права на участие в управлении) торгового портфеля и производные финансовые инструменты под влиянием факторов, связанных как с эмитентом фондовых ценностей и производных финансовых инструментов, так и общими колебаниями рыночных цен на финансовые инструменты.

Валютный риск - риск убытков вследствие неблагоприятного изменения курсов иностранных валют и (или) драгоценных металлов по открытым кредитной организацией позициям в иностранных валютах и (или) драгоценных металлах.

Процентный риск - риск возникновения финансовых потерь (убытков) вследствие неблагоприятного изменения процентных ставок по активам, пассивам и внебалансовым инструментам кредитной организации.

Риск ликвидности - риск убытков вследствие неспособности кредитной организации обеспечить исполнение своих обязательств в полном объеме. Риск ликвидности возникает в результате несбалансированности финансовых активов и финансовых обязательств кредитной организации (в том числе вследствие несвоевременного исполнения финансовых обязательств одним или несколькими контрагентами кредитной организации) и (или) возникновения непредвиденной необходимости немедленного и единовременного исполнения кредитной организацией своих финансовых обязательств.

Операционный риск - риск возникновения убытков в результате несоответствия характеру и масштабам деятельности кредитной организации и (или) требованиям действующего законодательства внутренних порядков и процедур проведения банковских операций и других сделок, их нарушения служащими кредитной организации и (или) иными лицами (вследствие

некомпетентности, непреднамеренных или умышленных действий или бездействия), несоразмерности (недостаточности) функциональных возможностей (характеристик) применяемых кредитной организацией информационных, технологических и других систем и (или) их отказов (нарушений функционирования), а также в результате воздействия внешних событий.

Правовой риск - риск возникновения у кредитной организации убытков вследствие:

- несоблюдения кредитной организацией требований нормативных правовых актов и заключенных договоров;
- допускаемых правовых ошибок при осуществлении деятельности (неправильные юридические консультации или неверное составление документов, в том числе при рассмотрении спорных вопросов в судебных органах);
- несовершенства правовой системы (противоречивость законодательства, отсутствие правовых норм по регулированию отдельных вопросов, возникающих в процессе деятельности кредитной организации);
- нарушения контрагентами нормативных правовых актов, а также условий заключенных договоров.

Риск потери деловой репутации кредитной организации (репутационный риск) - риск возникновения у кредитной организации убытков в результате уменьшения числа клиентов (контрагентов) вследствие формирования в обществе негативного представления о финансовой устойчивости кредитной организации, качестве оказываемых ею услуг или характере деятельности в целом.

Стратегический риск - риск возникновения у кредитной организации убытков в результате ошибок (недостатков), допущенных при принятии решений, определяющих стратегию деятельности и развития кредитной организации (стратегическое управление) и выражающихся в неучете или недостаточном учете возможных опасностей, которые могут угрожать

деятельности кредитной организации, неправильном или недостаточно обоснованном определении перспективных направлений деятельности, в которых кредитная организация может достичь преимущества перед конкурентами, отсутствии или обеспечении в неполном объеме необходимых ресурсов (финансовых, материально-технических, людских) и организационных мер (управленческих решений), которые должны обеспечить достижение стратегических целей деятельности кредитной организации.

Интересным будет рассмотреть классификацию рисков, приведенных в работе «The practice of Risk Management» [134].

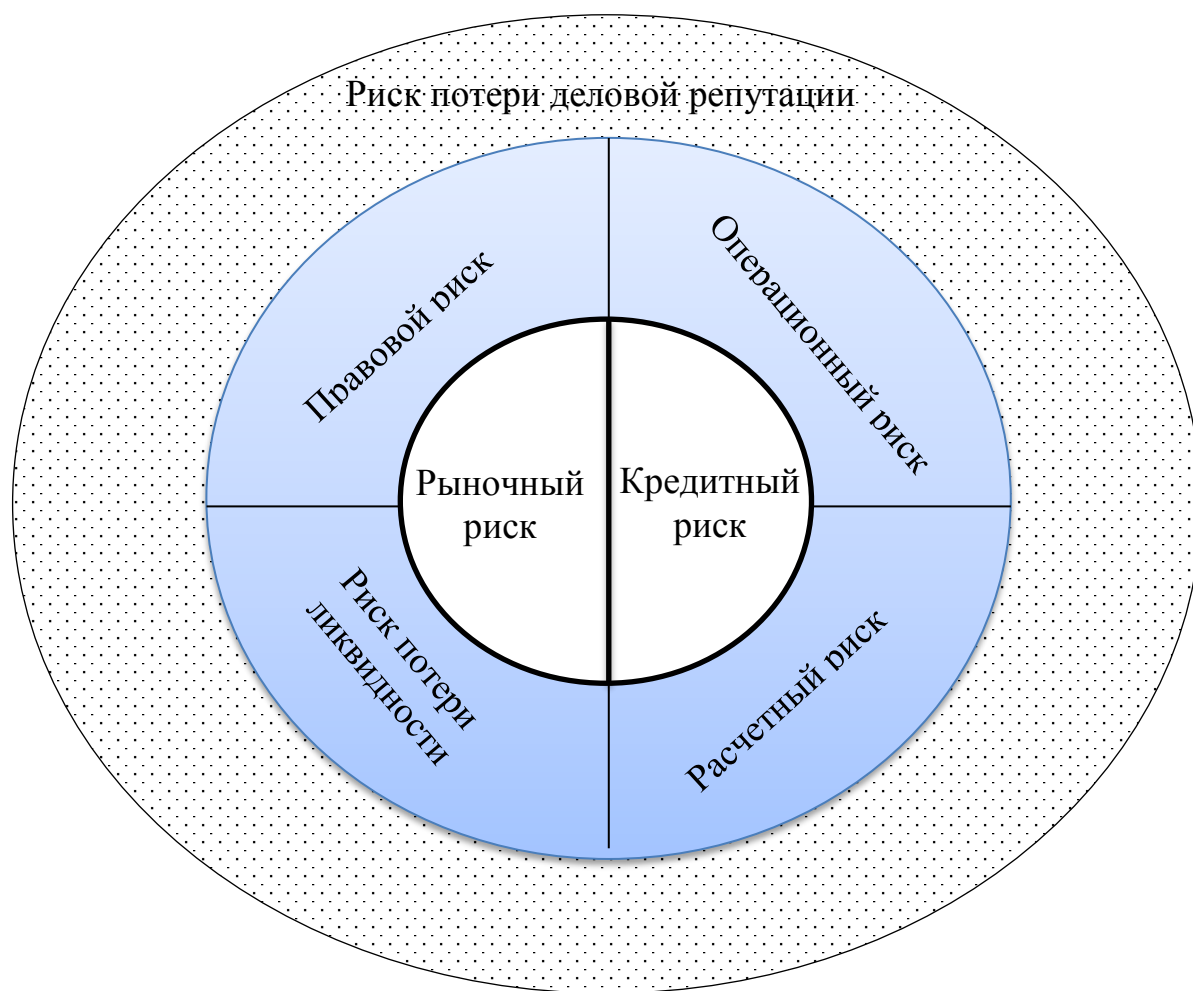


Рис. 1. Классификация рисков

Риски на рисунке 1 распределены по степени возможности управления и прогнозирования банком данных рисков. Рисками, расположенными в центре схемы, являются кредитный и рыночный риски, связанные с основными направлениями банковской деятельности — финансовыми вложениями и

кредитованием. Данные риски являются наиболее управляемыми, поскольку кредитные организации имеют возможность оценить риски финансовых вложений и кредитные риски до заключения сделки, так же указанные риски могут быть оценены с высоким уровнем доверительного интервала. Правовой, операционный риск, расчетный риск и риск потери ликвидности *возникают* при осуществлении деятельности кредитной организации, возможность управления банком данными рисками значительно меньше, чем рисками, расположенными в центре. Отмечаем, что данные риски могут нанести ущерб кредитной организаций в не меньшем размере чем кредитный и рыночный риски. Так, британский банк HSBC Holdings Plc в декабре 2012 года выплатил штраф в размере 1,5 млрд. долларов США за нарушение режима санкций, что является примером правового риска. Заметим, что банк HSBC признал, что не обладает достаточно надежными инструментами контроля над подобной деятельностью.

В письме Банка России от 23 июня 2004 г №70-Т "О типичных банковских рисках" отдельно не выделяется расчетный риск. Расчетный риск связан с неисполнением или задержкой исполнения транзакции в системе исполнения сделок. В целом данный риск можно рассматривать как частный случай операционного риска. Отдельное выделение данного риска в работе «The practice of Risk Management» и современной западной банковской практике связано с тем, что данный риск возникает у кредитной организации вне зависимости от качества операционных процедур банка. В работе Херштатский риск [69] рассмотрен один из примеров возникновения расчетного риска.

На момент закрытия (26 июня 1974 года) банка Херштатт (нем. Herstatt-Bank) органами надзора за банковской системой Германии, банк имел неурегулированные расчеты по сделкам валюте на сумму 200 млн. долларов. Закрытие данной кредитной организации повлекло за собой цепочку не платежей, что в свою очередь повлекло за собой снижение оборота платежной системы США на 60%.

На текущий момент банкротство банка Херштатт было пока единственным случаем реализации расчетного риска такого масштаба. Но меньшие по размерам

случаи расчетного риска периодически происходят. Так например, банкротство Drexel Burnham Lambert Trading в феврале 1990 года вызвало проблемы на рынке FOREX и рынке золота. Ликвидация в 1995 году банка Baring Brothers привела к сбоям в работе ряда европейских клиринговых систем [69]. В работе 56 на примере предприятия рассмотрена методика диверсификация рисков при работе с внешними контрагентами, данная методика также может быть применена для минимизации расчетных рисков [56].

Риск потери деловой репутации является наиболее сложно контролируемым и прогнозируемым риском, при этом данный риск может нанести кредитной организации наиболее серьезные потери и повлечь за собой банкротство кредитной организации. Наиболее яркими примерами реализации риска потери деловой репутации являются случаи банковской паники, когда в связи с утратой доверия к банковской системе в целом вкладчиками осуществляется массовое изъятие депозитов, а кредитные организации закрывают лимиты на рынке межбанковского кредитования. Недавним примером банковской паники в России является банковская паника в Самаре в 2013, когда из-за действий Центрального Банка России по отзыву лицензий у кредитных организаций было утрачено доверие к региональным банкам, что повлекло за собой массовое изъятие денежных средств вкладчиками, из-за чего Самарский банк "Солидарность" вынужден был пройти процедуру санации.

#### *1.1.2. Капитал банка как основной источник поглощения рисков.*

Капитал банка – собственные средства банка. Капитал банка первоначальный источник средств банка, поддерживающий доверие клиентов и кредиторов к банку. Согласно базельским соглашениям собственные средства (капитал) банка должны превышать 10% от взвешенных по риску активов, с учетом операционных, рыночных и иных рисков, так же кредитной организацией должны быть созданы буферы капитала, накапливаемые в период роста экономики, которые затем могут быть использованы в период рецессии. Чем выше соотношение капитала к активам банка, тем более устойчивым принято



считать банк. Показатель достаточности капитала отнесен к числу важнейших показателей при оценке надежности банка [1,2].

Как указывается в работах Симановского А.Ю. [71-76,97], главная функция капитала заключается в поглощении возможных убытков в случае реализации рисков, что должно обеспечивать защиту интересов вкладчиков и кредиторов банка. В случае реализации рисков (кредитного, операционного, рыночного и иных) размер капитала банка должен быть достаточен для покрытия (поглощения) ущерба, то есть риск вкладчиков и кредиторов банка должен быть минимален. Основной пруденциальный надзор направлен на определение достаточности капитала банка к погашению возможного ущерба (рисков). В случае если капитала банка недостаточно для погашения возможных рисков, то регулируемыми органами принимается решение о санации (финансовом оздоровлении) или кредитной организации [71-76] .

В международной практике расчет капитала банка регулируется рекомендациями Базельского Комитета, основанного в 1974 году при Банке международных расчетов. В него входят Центральные Банки крупнейших государств. Комитет разрабатывает рекомендации и стандарты Банковского надзора, применяемые органами банковского регулирования и надзора разных стран.

Крупные банкротства кредитных организаций, фондов и институциональных инвесторов, наблюдавшиеся в 70-80-е годы, «потребовали» от надзорных органов и банковского сообщества разработки соглашений, которые позволили бы защитить интересы кредиторов. В 1988 году Базельским Комитетом по банковскому надзору было разработано Соглашение по достаточности капитала (Базель I).

Изначально данное соглашение имело рекомендательный характер, однако в дальнейшем становится обязательной нормой (более 100 стран полностью или частично присоединились к Базелю I).

Базельским комитетом было предложено разделить капитал на 2 уровня в зависимости от источников капитала.

1 уровень (основной капитал банка) включал в себя наиболее качественные источники капитала: объявленный акционерный капитал, эмиссионный доход, нераспределенную прибыль прошлых лет, обязательные резервы.

2 уровень (дополнительный капитал) – менее качественные источники, в том числе переоценка активов, гибридные инструменты (привилегированные акции и субординированные кредиты).

Предполагалось, что капитал 2 уровня не должен превышать капитал 1 уровня.

До 1996 года достаточность капитала определялась по следующей формуле:

$$\frac{K1 + K2}{ApCredit} \geq 8\% \quad (1)$$

где K1 и K2 – капитал первого и второго уровня соответственно, а ApCredit – взвешенные по риску активы.

В дальнейшем в знаменатель дроби было принято решение включить операционные и рыночные риски.

В рамках Базеля I взвешенные по риску активы определялись исходя из коэффициентов риска актива, определенных Базельским Комитетом. В Базеле II («Международная конвергенция измерения капитала и стандартов капитала: новые подходы») кредитным организациям была предоставлена возможность определять размер взвешенных по риску активов и рыночных рисков исходя из внутренних рейтинговых моделей.

Основываясь на уроках финансового кризиса 2008–2009 гг., Базельский комитет по банковскому надзору (BCBS) приступил к пересмотру своих требований к достаточности капитала банков. Результатом стала система требований к достаточности капитала и ликвидности, получившая название Базель III (Basel III) и одобренная «Большой двадцаткой» на саммите в Сеуле в ноябре 2010 г. [67]

В рамках Базеля III были ужесточены требования к основному капиталу банка, введен капитал 3 уровня, введены буферы капитала призванные увеличить

достаточность капитала системообразующих банков, введены коэффициенты краткосрочной ликвидности (Liquidity Coverage Ratio) (LCR) и чистого стабильного фондирования Net Stable Funding Ratio (NSFR).

С 2014 года в Российской Федерации вступили в силу требования Положения Банка России от 28 декабря 2012 г. № 395-П «О методике определения величины собственных средств (капитала) кредитных организаций ("Базель III")», в соответствии с которыми капитал разделяется на 3 уровня: основной капитал, дополнительный капитал и добавочный капитал [4].

К основному капиталу, в соответствии с данным положением, отнесены наиболее «качественные» источники капитала: уставный капитал, эмиссионный доход, нераспределенная прибыль кредитной организации, подтвержденная аудиторской организацией. Из капитала вычитаются нематериальные активы и вложения в финансовые организации. В соответствии с инструкцией Банка России №139-И [5] достаточность основного капитала банка должна составлять не менее 5%.

В состав добавочного капитала включаются менее качественные источники капитала банка, такие как привилегированные акции кредитных организаций, нераспределенная прибыль текущего года, неподтвержденная аудиторской организацией и субординированные кредиты. Из капитала исключаются ненадлежащие активы (вложения в акции финансовых организаций, вложения в основные средства и прочие), превышающие определенные пороговые значения в зависимости от объема основного капитала банка. В соответствии с инструкцией Банка России №139-И сумма основного и дополнительного капитала банка должна составлять не менее 5,5% (6% с 01.01.2015).

К дополнительному капиталу относятся наименее качественные активы кредитных организаций, такие как субординированные кредиты, неудовлетворяющие условиям положения Банка России №395-П.

### *1.1.3. Взвешенные по риску активы. Стандартизированный подход и подход на основе внутренних кредитных рейтингов (IRB approach)*

С целью определения достаточности капитала кредитной организации активы кредитной организации, подверженные кредитному риску, взвешиваются исходя из уровня рисков данных активов [80,81,88,91,118,120,145,162,163,164].

В рамках Международной конвергенции измерения капитала и стандартов капитала: новые подходы (Базель II) для оценки взвешенных по риску активов предусмотрены 2 подхода – стандартизированный и подход на основе внутренних рейтингов (Internal Ratings-Based Approach (IRB Approach) [139,141].

Стандартизированный подход основан на взвешивании величины кредитных требований на коэффициент (коэффициент Кука), присваиваемый тому или иному заемщику в зависимости от внешнего кредитного рейтинга, определенного тем или иным международным рейтинговым агентством.

Второй подход, являющийся наиболее значимым изменением по сравнению с Базелем I, состоит в принципиальном допуске внутренних рейтингов к определению кредитоспособности заемщиков (подход Internal Rated Based Approach — IRB Approach). С точки зрения измерения кредитного риска подход IRB представляет собой математическую модель, учитывающую четыре фактора: вероятность дефолта контрагента (PD); удельный вес потерь при дефолте контрагента (LGD); абсолютную величину потерь при дефолте (EAD) и остаточный срок кредита или обращения долговой ценной бумаги (M). С применением этих показателей определяются так называемые ожидаемые (EL) и неожиданные (UL) потери, величина которых включается в расчет достаточности капитала [71-72].

Подробно основы стандартизированного подхода рассмотрены в работе О.А. Зайцевой [26]. В 3 последующих абзацах приведены цитаты из данной работы.

При существующей шкале внешних рейтингов от AAA до C– Базель II принимает во внимание рейтинги не ниже B–. Юридическим лицам с более

низкими рейтингами (независимо от их отраслевой принадлежности) присваивается повышенный коэффициент взвешивания (в размере 150%).

Базелем II выделяются рейтинги, присвоенные государствам и центральным банкам (так называемый рейтинг суверена), индивидуальные кредитные рейтинги того или иного предприятия или кредитной организации и экспортные кредитные рейтинги [28].

Степень риска контрагентов определяется следующим образом:

- в отношении суверенов — в зависимости от внешнего кредитного рейтинга государства;
- в отношении местных органов власти, а также банков предусмотрено два способа. Первый основан на внешнем кредитном рейтинге государства, резидентом которого является тот или иной банк или местный орган власти. Второй — на рейтинге самого банка или местного органа власти;
- в отношении юридических лиц (кроме банков) — в зависимости от рейтинга юридического лица;
- в отношении компаний, осуществляющих операции с ценными бумагами (профессиональные участники рынка ценных бумаг), выбирается система взвешивания в зависимости от надзорных принципов и методов регулирования их деятельности. В случае если регулирование деятельности профессиональных участников рынка ценных бумаг осуществляется в полном соответствии с требованиями Базеля II, то может быть применен подход, аналогичный банкам. Относительно российских условий это означает, что действующие правила надзора за участниками фондового рынка пока не в полной мере соответствуют стандартам Базельского комитета и к ним необходимо применять коэффициенты взвешивания, установленные для юридических лиц.

Для каждого из контрагентов, исходя из их внешних рейтингов, установлены соответствующие коэффициенты риска, так например кредит,

выданный юридическому лицу с рейтингом A+, установленным рейтинговым агентством Standard & Poor's, коэффициент риска равняется 50%, а для кредита, выданному юридическому лицу с рейтингом ниже B-, коэффициент риска равняется 150%. Приведем пример расчета взвешенных по риску активов, так если обоим юридическим лицам выдан кредит на сумму 100 млн. руб., то взвешенные по риску активы по кредиту юридическому лицу с рейтингом A+ будут равняться 50 млн. руб. ( $100 \cdot 50\%$ ), а для юридического лица с рейтингом B- взвешенные по риску активы составят 150 млн. руб. ( $100 \cdot 150\%$ ). Соответственно, для выдачи кредита юридическому лицу с рейтингом A+ по формуле (1) потребуется как минимум 4 млн. руб. капитала банка ( $50 \cdot 8\%$ ), тогда как для кредита юридическому лицу с рейтингом B- потребуется 12 млн. руб. капитала банка ( $150 \cdot 8\%$ ). То есть, более рискованные кредиты требуют значительно большего объема капитала для покрытия риска [26].

Заметим, что коэффициенты кредитного риска, определенные Базельским Комитетом, установлены исходя из экспертных оценок уровня кредитных требований, не подкреплены достаточными статистическими исследованиями, так же имеются сомнения к качеству внешних рейтингов, в связи с проциклическостью рейтингов и к тому обстоятельству, что аутсорсинг надзорных оценок выступает для рейтинговых агентств дополнительным фактором конфликта интересов между репутацией (объективность оценок) и доходом (торговля рейтингами) [128,129]. В условиях, когда функционирование рейтинговой индустрии контролируется только «рынком», весьма вероятным является разрешение упомянутого конфликта не в пользу объективности оценок

Перейдем к расчету взвешенных по риску активов на основе внутренних кредитных рейтингов (IRB подхода).

IRB подход, представленный в Базель II, являлся дальнейшим развитием идей оценки рисков на основе вероятности неисполнения обязательств заемщиков, заложенных в Базель I. IRB подход идеологически заимствован у наиболее продвинутых банков, проводивших расчеты так называемого

экономического капитала с 80-х годов прошлого века, но с исправлениями и дополнениями, ориентированными на консервативные традиции надзора.

IRB-подход к оценке базовых требований по капиталу исходит из следующих методических посылок [27,76]:

- премия за риск ориентирована на средние или близкие к средним потерям, т. е. на средние значения вероятности дефолта заемщика (PD) и уровень потерь в случае дефолта (LGD) в разрезе продуктов и групп клиентов;
- фактические потери могут значительно превышать средние ожидаемые значения. Неожидаемые потери в общем случае должен покрывать капитал;
- вероятность потерь может быть описана функцией нормального распределения, следовательно, уровень потерь, которые должны быть покрыты капиталом, задается принятым уровнем надежности.

Внедрение подхода к оценке на основе внутренних кредитных рейтингов является наиболее существенной новацией Базеля II, поскольку кредитным организациям позволено самостоятельно на основе собственной статистике о случаях неисполнения заемщиками кредитных обязательств (дефолтов) определять уровни кредитного риска.

IRB подход подразделяется на базовый IRB подход и продвинутый IRB подход (advance IRB approach). В рамках базового IRB подхода кредитные организации самостоятельно определяют вероятность дефолта (PD) по заемщикам, которые распределены по группам (рейтингам) в зависимости от уровня рисков, и величина кредитного требования, подверженная риску дефолта (EAD), тогда как уровень потерь в случае дефолта (LGD) определяется регулятором.

В случае применения продвинутого IRB подхода предполагается, что кредитные организации самостоятельно определяют уровень потерь в случае дефолта (LGD) [76].

Предполагается, что самостоятельный расчет рисков кредитными организациями по IRB подходу позволит корректнее учитывать риски при расчете капитала, поскольку кредитные организации обладают наиболее полной базой о

реализованных рисках и потерях в случае реализации рисков, так же кредитные организации заинтересованы в корректной оценке принимаемых рисков. В работе Банковское регулирование: революция [72,73] отмечается, что введенный Базелем I подход к оценке достаточности капитала с учетом взвешивания активов на риск (коэффициент Кука) явился стимулом к бурному развитию секьюритизации активов, что послужило одной из основных предпосылок финансового кризиса 2008 года.

Рассмотрим подробнее методику расчета взвешенных по риску активов в соответствии с IRB подходом. Банком России опубликовано письмо от 29 декабря 2012 г. № 192-Т "О Методических рекомендациях по реализации подхода к расчету кредитного риска на основе внутренних рейтингов банков", в котором приведены формулы расчета взвешенных по риску активов для применения IRB подхода для кредитных организаций Российской Федерации [6].

Для расчета взвешенных по риску кредитных требований используются следующие компоненты кредитного риска:

- вероятность дефолта (PD, %) - числовое значение вероятности дефолта заемщика (контрагента) по конкретному кредитному требованию (пулу однородных кредитных требований) на период в один год, соответствующее разряду внутренней рейтинговой шкалы заемщиков (пула однородных заемщиков), к которому отнесено кредитное требование (пул однородных кредитных требований). Минимально возможное значение вероятности дефолта по кредитным требованиям, учтенным в классах кредитных требований к корпоративным заемщикам, кредитных требований к финансовым институтам и кредитных требований к розничным заемщикам составляет 0,03%. Значение вероятности дефолта по заемщикам (контрагентам), находящимся в состоянии дефолта, составляет 100%; уровень потерь при дефолте (LGD, %) - доля безвозвратных потерь при дефолте в величине кредитного требования к контрагенту по конкретному кредитному требованию (пулу однородных кредитных требований);



- величина кредитного требования, подверженная риску дефолта на момент возможного дефолта (EAD, рублей), - средства, предоставленные заемщику и не погашенные им, включая комиссии, штрафы и недополученные проценты;

- срок до погашения кредитного требования (M, лет);

- конверсионный коэффициент (CCF, %) предназначен для расчета кредитного эквивалента условного обязательства кредитного характера путем умножения конверсионного коэффициента на сумму условного обязательства кредитного характера.

Компоненты кредитного риска (PD, LGD, EAD, M) включаются в формулы, используемые для расчета величины кредитного риска по каждому кредитному требованию (в случае кредитных требований к розничным заемщикам и приобретенных прав кредитных требований - пулу однородных кредитных требований). Компоненты кредитного риска (PD, LGD, EAD) используются также для расчета величины ожидаемых потерь в целях определения регулятивного капитала банка.

Расчет величины взвешенных по риску кредитных требований, по которым не произошел дефолт ( $PD \neq 100\%$ ), для кредитных требований к корпоративным, суверенным заемщикам и финансовым институтам, рассчитывается по следующей формуле:

$$RWA = \alpha * 12,5 * EAD * LGD * \left( N \left( \frac{N^{-1}(PD) + \sqrt{R} * N^{-1}(0,999)}{\sqrt{1-R}} \right) - PD \right) * \frac{1 + (M - 2,5) * b(PD)}{1 - 1,5 * b(PD)} \quad (2)$$

где R - значение показателя корреляции, рассчитываемое по формуле:

$$R(PD) = 0,2 * \frac{1 - e^{-35*PD}}{1 - e^{-35}} + 0,24 * \left( 1 - \frac{1 - e^{-35*PD}}{1 - e^{-35}} \right). \quad (3)$$

b(PD) - значение показателя корректировки на срок до погашения:

$$b(PD) = (0,11852 - 0,05478 * \ln(PD))^2 \quad (4)$$

N(x) - стандартное нормальное распределение;

Для кредитных требований к розничным заемщикам применяется продвинутый IRB подход, то есть банк самостоятельно определяет значения компонентов кредитного риска (PD, LGD, EAD).

Расчет величины взвешенных по риску кредитных требований осуществляется с помощью следующей формулы:

$$RWA = 12,5 * EAD * LGD * \left( N \left( \frac{N^{-1}(PD) + \sqrt{R} * N^{-1}(0,999)}{\sqrt{1-R}} \right) - PD \right) \quad (5)$$

где R - показатель корреляции, значение которого установлено равным:

0,04 - для активов, отнесенных к подклассу "возобновляемые розничные кредитные требования";

0,15 - для активов, отнесенных к подклассу "кредитные требования, обеспеченные залогом жилой недвижимости".

Значение показателя корреляции для требований, отнесенных к подклассу "прочие розничные кредитные требования", рассчитывается по формуле:

$$R(PD) = 0,03 * \frac{1 - e^{-35*PD}}{1 - e^{-35}} + 0,16 * \left( 1 - \frac{1 - e^{-35*PD}}{1 - e^{-35}} \right) \quad (6)$$

Приведем пример того, как изменяются взвешенные по риску активы, отнесенные к подклассу "кредитные требования, обеспеченные залогом жилой недвижимости" в зависимости от изменения вероятности дефолта.

Пусть EAD равно 100 млн. руб., LGD равняется 50%, R равняется 0,15, тогда при изменении PD взвешенные по риску активы будут изменяться так:

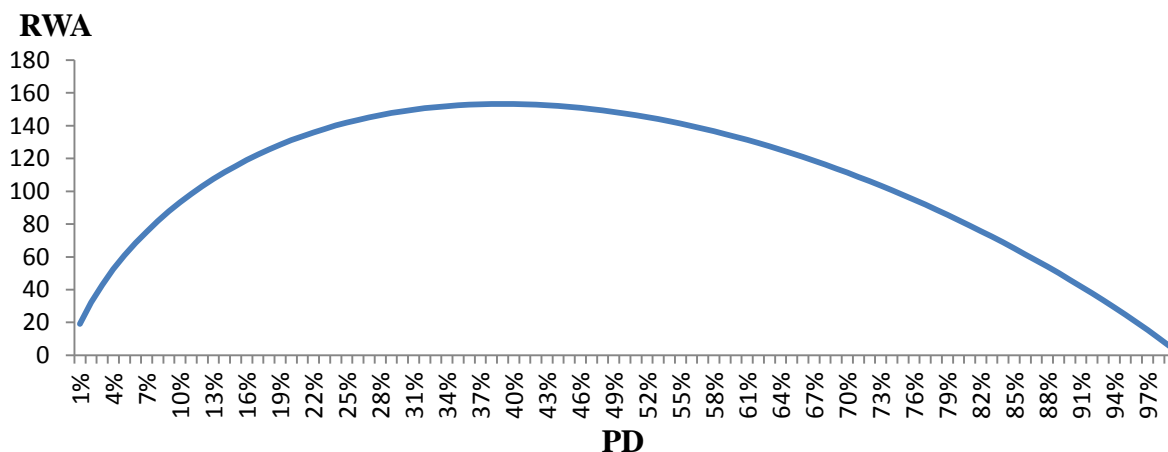


Рис. 2. Изменение взвешенных по риску активов (RWA) при изменении вероятности дефолта (PD).

Из графика, приведенного на рисунке 2, видно что взвешенные по риску активы сначала возрастают с увеличением вероятности дефолта (PD), а затем начинают снижаться. Может показаться некорректным, что взвешенные по риску активы начинают снижаться после достижения вероятности дефолта определенного порогового значения. Снижение взвешенных по риску активов объясняется тем, что в рамках IRB подхода Базеля II предполагается, что резервы, формируемые кредитной организацией, не должны быть меньше ожидаемых потерь по портфелю, иначе разница между сформированными резервами и ожидаемыми потерями должна вычитаться из капитала банка (§43, Базель II [2]). Следовательно, часть кредитных рисков актива покрывается формируемыми кредитной организацией резервами.

Заметим, что, несмотря на несомненные выгоды от перехода кредитных организаций к оценке капитала на основе IRB подхода, авторами отмечаются определенные минусы.

Так, гипотеза, что кредитный риск по продукту или группе клиентов определяется через средние значения PD и LGD, недостаточно подтверждена эмпирическими исследованиями. Заметим, что в рамках IRB подхода не учитывается волатильность риска по конкретному заемщику и группе, в связи с чем, уровень риска может значительно превысить средние значения [60].

Распределение вероятности дефолтов (PD) задается функцией нормального распределения, тогда как уровень дефолтов в случае значительных стрессов скорее описывается функциями из области теории катастроф. Заметим, что аргументом использования нормального распределения может являться низкий уровень предсказуемости основных параметров риска (PD и LGD) в условиях системного стресса, и высокий уровень доверительного интервала в 99,9%.

## **1.2. Методы оценки кредитного риска**

В данной части рассмотрены основные методы оценки кредитного риска, которые можно разделить на количественные методы оценки риска (Value at Risk

(VaR) и матрицы переходов / миграций) и классификационные методы (скоринг, дерево принятия решений, дискриминационные методы). Такие методы как VaR и матрица переходов миграций чаще всего используются в тех случаях, когда необходимо оценить (спрогнозировать) потери кредитной организации в целом по портфелю / сегменту однородных кредитов, тогда как классификационные методы используются для оценки риска по конкретному заемщику [25,50,62,63,113,114,115,116,117,123,124,125,126,127,146,147,158,159,161,188].

### *1.2.1. Оценка кредитного риска по VaR методу*

VaR метод является универсальной методикой для оценки различных рисков (валютного, кредитного, ликвидности, операционного и других). Основной задачей данного метода является оценка максимальных потерь от риска в течение определенного периода времени с заданным доверительным интервалом [95,122,130,135,136,181]

Основными подходами к оценке VaR являются: исторический, когда на основе статистических данных об уровне потерь оцениваются потери с требуемым доверительным интервалом, параметрический, когда по статистическим данным определяется функция распределения вероятности возможных потерь.

При оценке кредитных рисков чаще всего VaR метод применяется для оценки максимальных потерь по портфелю за определенный период времени, поскольку оценка VaR по отдельному кредитному требованию не достаточна для корректной оценки риска актива, поскольку при расчете VaR не используются параметры позволяющие характеризовать заемщика.

Широкое применение VaR метод нашел при расчете рентабельности капитала, скорректированной на риск (RAROC) [78]. Показатель RAROC определяется как отношение чистой прибыли за вычетом ожидаемых вследствие экономического риска потерь к капиталу.

### 1.2.2. Применение матриц переходов / миграции к оценке кредитного риска

Одним из методов, позволяющих оценить вероятность дефолта (перехода на просрочку свыше 90 дней) по кредитному портфелю, являются матрицы миграций кредитов / ссудной задолженности [59,177].

Будем считать, что для каждого кредитного договора без просроченной задолженности на начало месяца существует только 2 возможных состояния на конец месяца – отсутствие просроченной задолженности по договору или наличие просрочки по договору. То есть, если 5,0% кредитных договоров вышли на просрочку, то у 95,0% не возникло просрочки. Соответственно для кредитных договоров с просроченной задолженностью от 1 до 30 дней на начало месяца существует 2 возможных состояния на конец месяца – отсутствие просроченной задолженности или наличие просроченной задолженности от 31 до 60 дней.

Матрица переходов/миграций представляет собой матрицу вероятностей перехода из одной группы просрочки в другую (размерность матрицы составляет  $n * p$ , где  $n$  – число групп просрочки,  $p$  – число групп возможных для перехода).

Далее в таблице приведен пример матрицы миграций.

Таблица 1. Матрица миграций

		Группа на конец месяца				
		Без просрочки	С просрочкой от 1 до 30 дней	С просрочкой от 31 до 60 дней	С просрочкой от 61 до 90 дней	С просрочкой свыше 90 дней
Группа на начало месяца	Без просрочки	95,0%	5,0%	0,0%	0,0%	0,0%
	С просрочкой от 1 до 30 дней	62,8%	0,0%	37,2%	0,0%	0,0%
	С просрочкой от 31 до 60 дней	30,6%	0,0%	0,0%	69,4%	0,0%
	С просрочкой от 61 до 90 дней	5,0%	0,0%	0,0%	0,0%	95,0%

Как видно из данной матрицы  $p > n$  поскольку оценивается вероятность потерь только для групп с просрочкой до 90 дней. В соответствии с подходами Базеля II кредиты с просрочкой свыше 90 дней признаются дефолтными.

Данную матрицу можно приставить в виде следующей схемы.

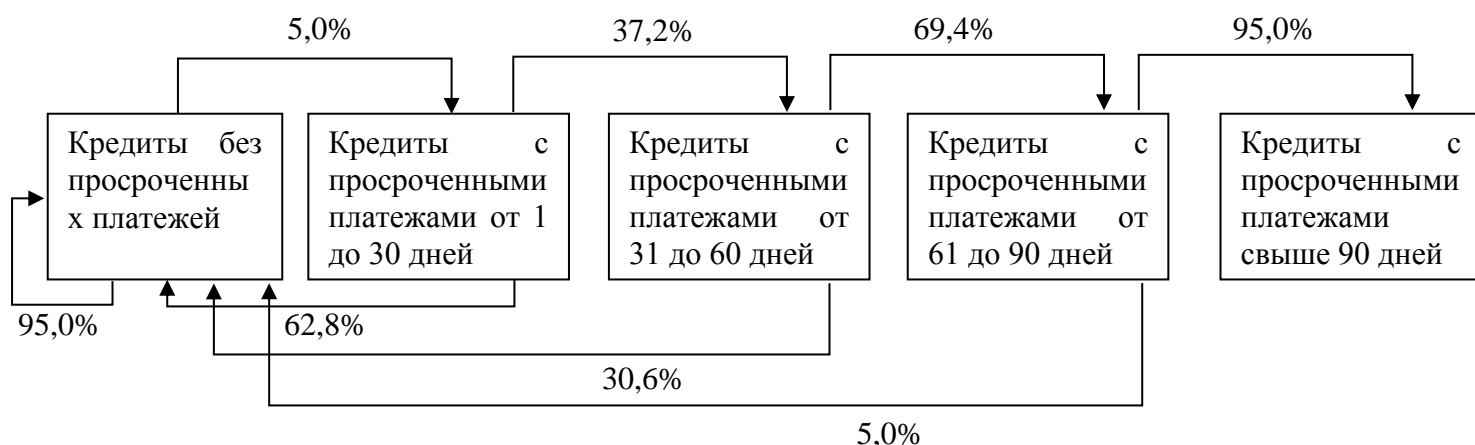


Рис. 3. Схема, иллюстрирующая матрицу переходов/миграций.

Согласно рис. 3, вероятность перехода ссуды за месяц из состояния «без просроченных платежей» в состояние «с просроченными платежами от 1 до 30 дней» составляет 5,0%, следовательно, с вероятностью 95,0% ссуда останется в состоянии «без просроченных платежей», частота перехода ссуды из состояния «с просроченными платежами от 1 до 30 дней» в состояние «с просроченными платежами от 31 до 60 дней» составляет 37,2%, тогда вероятность погашения задолженности (переход в состояние «без просроченных платежей») составляет 62,8%.

Исходя из матрицы миграций, можно построить динамическую модель, позволяющую рассчитать долю дефолтов кредитной организации за определенный период времени (объемы ссудной задолженности, перешедшие в дефолт за период / объемы ссудной задолженности без дефолта на начало периода).

На рисунке 4 дано графическое описание модели, используемой для расчета вероятности потерь банка.

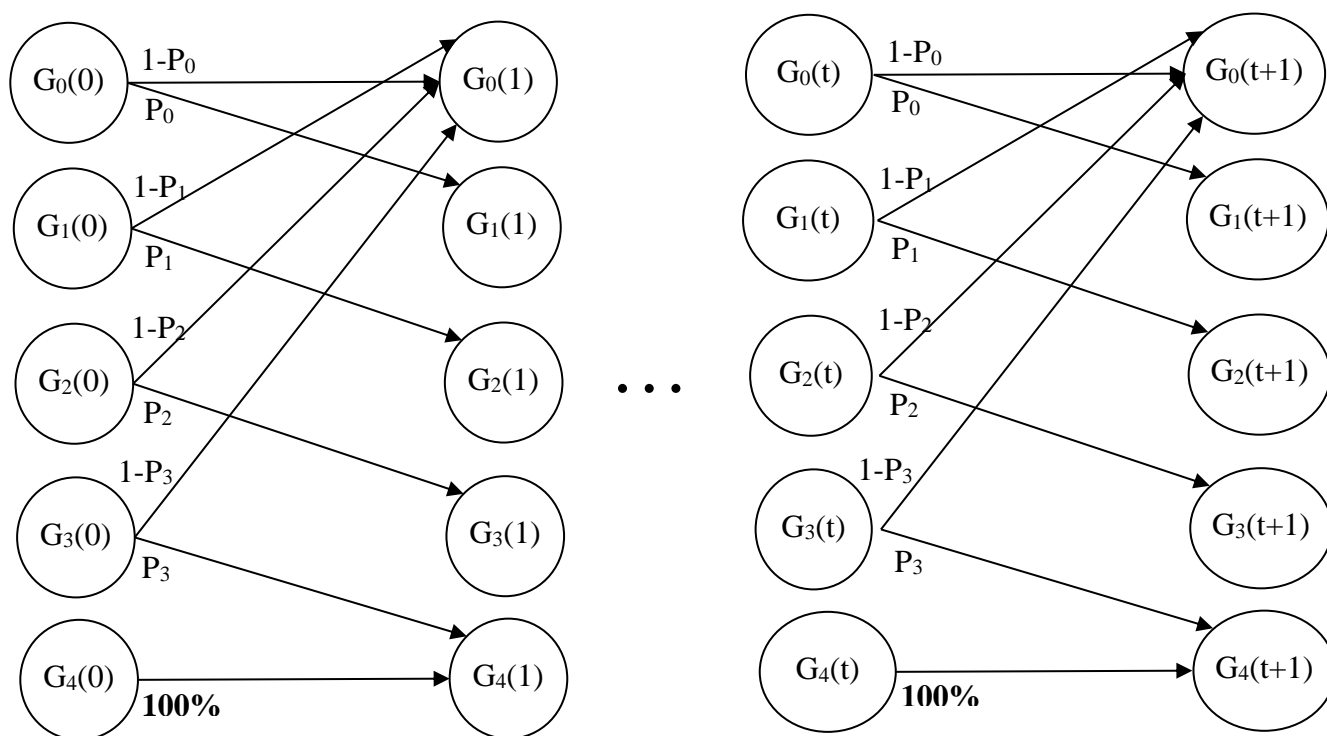


Рис. 4. Динамическая модель расчета вероятности дефолта

Константами модели выступают значения элементов матрицы переходов/миграций:

$P_0$  – вероятность перехода из группы без просроченных платежей в группу с просрочкой от 1 до 30 дней;

$P_1$  – вероятность перехода из группы с просрочкой от 1 до 30 дней в группу с просрочкой от 31 до 60 дней;

$P_2$  – вероятность перехода из группы с просрочкой от 31 до 60 дней в группу с просрочкой от 61 до 90 дней.

$P_3$  – вероятность перехода из группы с просрочкой от 61 до 90 дней в группу с просрочкой свыше 91 дня.

Динамическими переменными модели являются:

$G_0(t)$  – доля ссуд без просроченных платежей в периоде времени  $t$ ;

$G_1(t)$  – доля ссуд с просрочкой от 1 до 30 дней в периоде времени  $t$ ;

$G_2(t)$  – доля ссуд с просрочкой от 31 до 60 дней в периоде времени  $t$ ;

$G_3(t)$  – доля ссуд с просрочкой от 61 до 90 дней в периоде времени  $t$ ;

$G_4(t)$  – доля ссуд с просрочкой свыше 90 дней в периоде времени  $t$ ,

где  $t$  – период времени (месяц).

Для вычисления потерь для ссуд, находящихся в группе без просроченных платежей, в периоде  $t = 0$  устанавливаются следующие значения для групп:  $G_0(0) = 100\%$ ;  $G_1(0), G_2(0), G_3(0), G_4(0) = 0\%$ . Значения динамических переменных в периоде  $t+1$  вычисляются по следующим формулам:

$$\begin{aligned} G_0(t+1) &= G_0(t) * (1 - P_0) + G_1(t) * (1 - P_1) + G_2(t) * (1 - P_2) + G_3(t) * (1 - P_3); \\ G_1(t+1) &= G_0(t) * P_0; \\ G_2(t+1) &= G_1(t) * P_1; \\ G_3(t+1) &= G_2(t) * P_2; \\ G_4(t+1) &= G_4(t) + G_3(t) * P_3. \end{aligned} \quad (7)$$

Исходя из того, за какой период необходимо рассчитать потери по портфелю, определяется конечное число итераций модели.

Приведем пример расчета вероятности дефолта по портфелю в течение 12 месяцев.

Таблица 1. Расчет вероятности дефолта.

Срок просрочки на начало месяца	Период t					
	0	1	2	3	4	5
Без просрочки ( $G_0$ )	100,0%	95,0%	93,4%	92,2%	91,2%	90,1%
С просрочкой от 1 до 30 дней ( $G_1$ )	0,0%	5,0%	4,8%	4,7%	4,6%	4,6%
С просрочкой от 31 до 60 дней ( $G_2$ )	0,0%	0,0%	1,9%	1,8%	1,7%	1,7%
С просрочкой от 61 до 90 дней ( $G_3$ )	0,0%	0,0%	0,0%	1,3%	1,2%	1,2%
С просрочкой свыше 90 дней ( $G_4$ )	0,0%	0,0%	0,0%	0,0%	1,2%	2,4%

Срок просрочки на начало месяца	Период T						
	6	7	8	9	10	11	12
Без просрочки ( $G_0$ )	89,0%	88,0%	86,9%	85,9%	84,9%	83,9%	82,9%



С просрочкой от 1 до 30 дней ( $G_1$ )	4,5%	4,5%	4,4%	4,4%	4,3%	4,3%	4,2%
С просрочкой от 31 до 60 дней ( $G_2$ )	1,7%	1,7%	1,7%	1,6%	1,6%	1,6%	1,6%
С просрочкой от 61 до 90 дней ( $G_3$ )	1,2%	1,2%	1,2%	1,2%	1,1%	1,1%	1,1%
С просрочкой свыше 90 дней ( $G_4$ )	3,6%	4,7%	5,8%	6,9%	8,0%	9,1%	10,2%

Как видно из таблицы 1, вероятность дефолта (доля ссуд в дефолте) в течение 12 месяцев для группы без просроченных платежей равняется 10,2%.

Матрицы миграций применяются в том случае, когда необходимо спрогнозировать ожидаемые потери кредитной организации в течение определенного периода времени. Точность прогноза по матрицам миграции во многом зависит от волатильности вероятностей перехода из группы в группу. Заметим, что группами для матриц миграций могут служить не только группы, определенные исходя из срока просрочки, но и группы, определенные исходя из категорий качества актива [59].

### *1.2.3. Классификационные методы оценки кредитного риска*

Классификационные методы оценки кредитного риска, такие как кредитный скоринг, древо принятия решений, дискриминационные методы классификации, чаще всего применяются кредитной организацией для определения кредитного риска на момент принятия решения о выдаче кредита [96,103,105,106,107,110,182,185,186,187].

Классификационные методы основаны на том, что кредитной организацией формируется образы (классы) качественного заемщика и некачественного заемщика. Задача выделения образа заёмщика рассматривается в рамках теории распознавания образов (Pattern recognition) [121,148,149,150,151,153,154,155,166,170,171,172]

Образом (классом) – определенная группа объектов (наблюдений), выделенная на основании набора параметров, характеризующий объект (наблюдение).

Задача классификации (распознавания образа) – задача нахождения функции (набора функций), позволяющей на основании имеющихся параметров отнести объект (наблюдение) к требуемому классу (образу).

Рассмотрим подробнее классификацию заемщиков на основании скоринговой модели.

Кредитный скоринг — статистический прием, предложенный американским ученым Дэвидом Дюраном в 1941 г. [79,133,137,138,140,152,156,160,165,178], который первоначально использовался для классификации розничных заемщиков. При использовании скоринг модели каждому заемщику присваивается рейтинг, который характеризует его финансовое состояние и способность своевременно погасить свои обязательства перед кредитором. В дальнейшем весь диапазон возможных значений разбивается на интервалы, рейтинговые группы. Посредством калибровки модели на основе исторических данных каждому рейтинговому баллу ставится в соответствие вероятность дефолта, которая по большей части является определением доли заемщиков этой группы, испытавших дефолт в течение года.

На данный момент модели кредитного скоринга являются весьма распространенными в банковской практике, особенно для оценки вероятности дефолта по однородным портфелям розничных ссуд [179,180]. Однако первоначальная модель скоринговой оценки Дюрана была весьма упрощенной. Автор рассматривал определенные характеристики каждого розничного заемщика (такие как пол, возраст, срок проживания на одном месте, профессия, место и опыт работы, финансовое состояние) и в зависимости от их значения присваивал каждому клиенту определенный балл. Если суммарная оценка всех характеристик превышала 1,25 балла, то клиент считался кредитоспособным. [79]

Основным плюсом кредитного скоринга является относительная простота расчетов, поскольку чаще всего используются линейные зависимости, такие как

допустим рост доходов заемщика уменьшает скоринговый балл, а рост суммы кредитов увеличивает скоринговый балл [183,184].

Минусом скоринга является относительно более низкая точность классификации по сравнению с более продвинутыми методами классификации такими как древопринятия решения, Logit метод, нейронные сети [79,167,169].

Одним из часто применяемых методов для оценки кредитного риска является метод *древа* принятия решений [132,173,174].

Дерево принятия решений — один из методов принятия решений, использующийся в статистике и анализе данных. Структура дерева представляет собой «листья» и «ветки». На ребрах («ветках») дерева решения записаны параметры, от которых зависит вероятность возврата кредита, в «листьях» записаны значения вероятности возврата.

Представим вариант древа принятия решений для оценки кредитного риска.

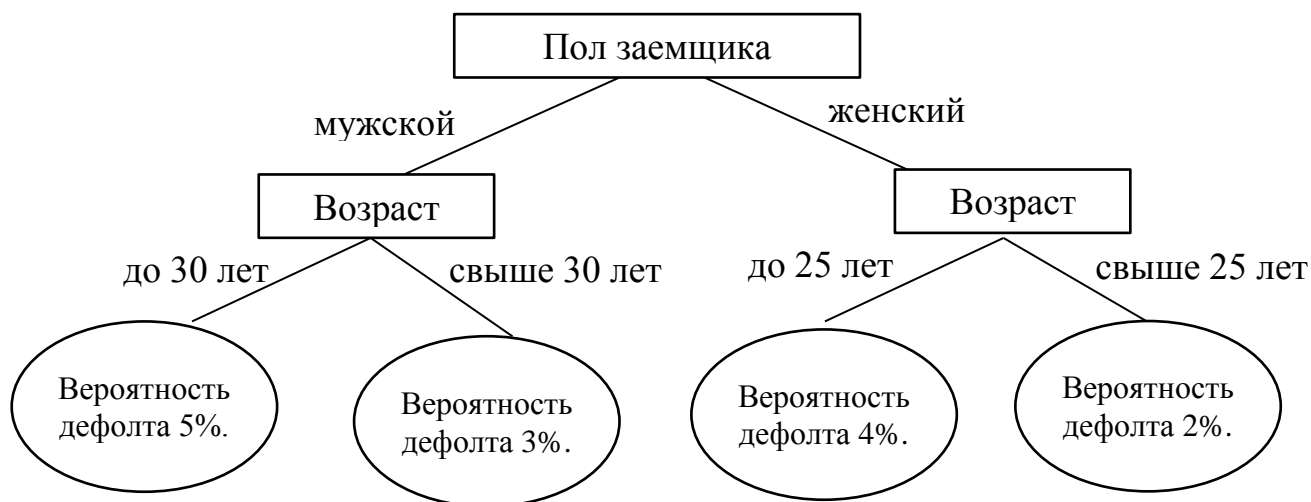


Рис. 5. Пример древа принятия решений для оценки риска заемщиков

Как видно из рисунка 4, чтобы классифицировать заемщика, надо спуститься по дереву до листа и выдать соответствующее значение. Цель древа принятия решений состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной на основе нескольких переменных на входе. [143].

Основными плюсами деревьев принятия решений являются:

1. Данный метод достаточно прост для понимания и интерпретации (используется модель «белого ящика», тогда как в моделях «черного ящика», как например нейронные сети, результаты данной модели поддаются объяснению с трудом);

2. Не требуется предварительной подготовки данных, многие методы классификации требуют предварительной нормализации параметров, введения фиктивных переменных, удаления пропущенных данных.

3. *Древо* принятия решений оперирует как с категориальными переменными (пол, должность и другие), так и с количественными переменными (доходы, сумма кредита, возраст и другие).

Недостатками метода являются:

1. Проблема получения оптимального дерева решений является NP-полной задачей с точки зрения некоторых аспектов оптимальности даже для простых задач [143, 168]. Практическое применение деревьев решений основано на эвристических алгоритмах, таких как алгоритм «жадности», которые не могут обеспечить оптимальность всего дерева в целом.

2. При использовании дерева принятия решений возникает проблема переобученности, поскольку при достаточно большом дереве на обучающей выборке можно со 100% точностью классифицировать заемщиков. Так, заемщик с суммой кредита от 100 до 101 тыс. руб. характеризуется как некачественный, а с суммой от 102 до 104 тыс. руб. характеризуется как качественный. В связи с чем возникает проблема определения оптимального размера дерева.

Далее рассмотрим методы классификации на основе построения линейного дискриминанта, такие как Logit и Probit модели [189,190].

Для оценки вероятности дефолта заемщиков банками применяются модели бинарного выбора, такие как Logit и Probit модели, в основе которых лежит построение линейного классификатора, позволяющего оценивать апостериорные вероятности принадлежности объектов классам. Основное различие в данных моделях является предположение о функции распределения вероятности

дефолта. В Probit модели предполагается, что функция вероятности описывается нормальным распределением, а в рамках Logit модели – логистической кривой.

Модель бинарного выбора включает два типа переменных:

- зависимую бинарную переменную вида  $y = [0; 1]$  принимающую значение 1, если по заемщику возник дефолт, и 0, если у заемщика отсутствуют случаи дефолта.
- совокупность параметров заемщика  $X$  (объясняющие переменные), таких как пол, возраст, сумма кредита и других.

Таким образом, вероятность дефолта  $i$ -го заемщика равняется  $P(y = 1) = F(X_i)$ , где  $F$ -функция стандартного нормального распределения, в случае Probit модели, или логистическая кривая в случае Logit модели.

Приведем графический пример оценки вероятности дефолта Logit моделью.

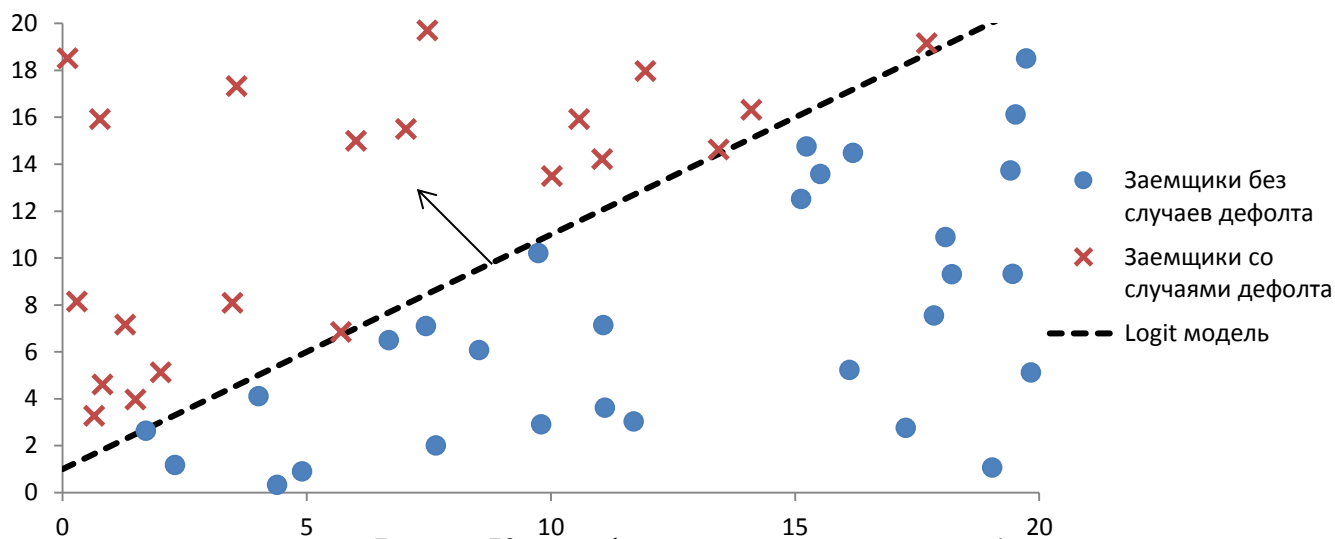


Рис. 6. Классификация заемщиков Logit моделью

Как видно из рисунка 6 заемщики со случаями дефолта расположены по одну сторону от линейной гиперплоскости, построенной с использованием Logit модели. С увеличением расстояния от гиперплоскости в направлении, указанного стрелкой на графике, возрастает вероятность дефолта, при увеличении расстояния в противоположную сторону вероятность дефолта уменьшается.

Основными плюсами данных методов являются:

1. Простота в построении Logit и Probit модели, что позволяет использовать значительные объемы данных о заемщиках при построении модели.

2. Однозначность интерпретации влияния параметров заемщика на вероятность дефолта, поскольку используется модель «белого ящика», так с возрастанием доходов заемщика Logit модель будет уменьшать вероятность дефолта, а с возрастанием суммы кредита вероятность дефолта будет увеличиваться.

Основными минусами данных методов являются:

1. Низкое качество классификации в случае, если классифицируемые множества не представляется возможным однозначно разделить линейной гиперплоскостью [61].

Так, например, для множества приведенного далее применение Logit или Probit моделей не представляется целесообразным.

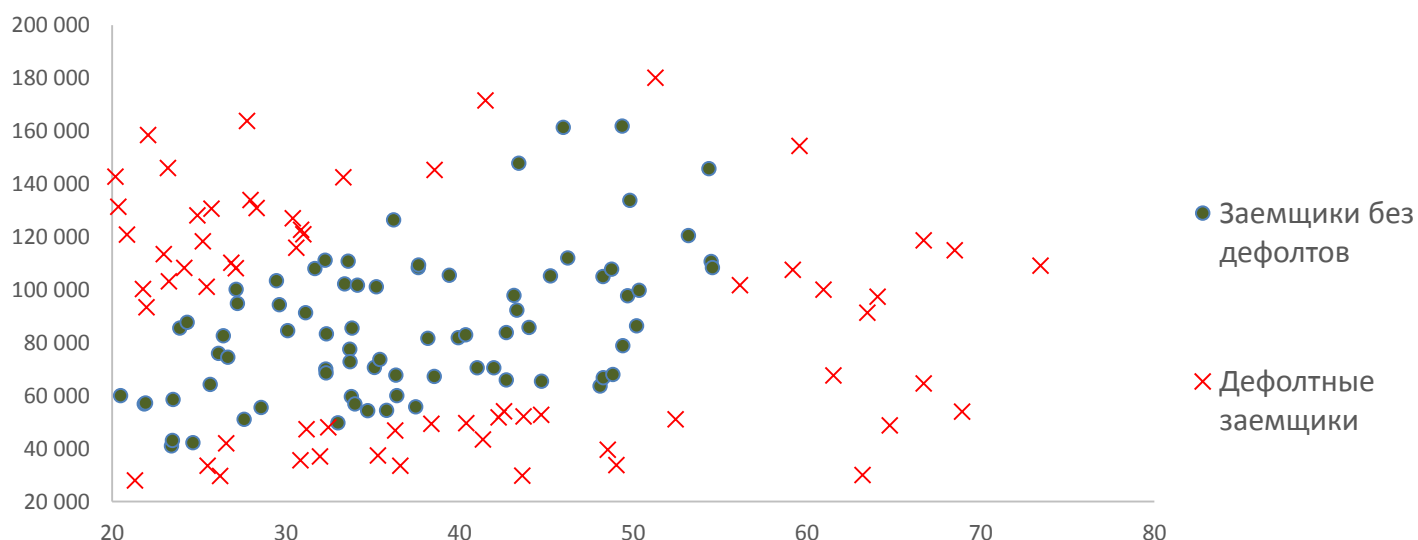


Рис. 7. Пример множества, неразделимого линейной гиперплоскостью.

2. Указанные модели очень чувствительны к мультиколлинеарности переменных. Так, если корреляция между доходом заемщика и суммой кредита значительна, например 90%, и при этом наблюдается снижение вероятности дефолта с ростом доходов, то Logit и Probit модели будут утверждать, что с ростом суммы кредита вероятность дефолта так же будет снижаться, что противоречит экономической логике [157].

Все методы, такие как кредитный скоринг, дерево принятия решений, Logit и Probit методы, являются моделями «белого ящика», что позволяет легко интерпретировать результаты данных моделей.

В заключении будут рассмотрены нейронные сети, которые являются модель «черного ящика», то есть интерпретация результатов данных моделей довольно затруднительна, не смотря на проблемы с интерпретацией, нейронные сети часто используются для оценки кредитоспособности заемщиков, поскольку имеют наиболее высокое качество классификации по сравнению с остальными методами [93,104,108,109,111,112,131,142,144].

Исторически первой работой, заложившей теоретический фундамент для создания искусственных моделей нейронов и нейронных сетей, принято считать опубликованную в 1943 г. статью Уоррена С. Маккалоха и Уальтера Питтса "Логическое исчисление идей, относящихся к нервной активности" [51]. Главный принцип теории Маккалоха и Питтса заключается в том, что произвольные явления, относящиеся к высшей нервной деятельности, могут быть проанализированы и поняты, как некоторая активность в сети, состоящей из логических элементов, принимающих только два состояния ("все или ничего") [15,18,19].

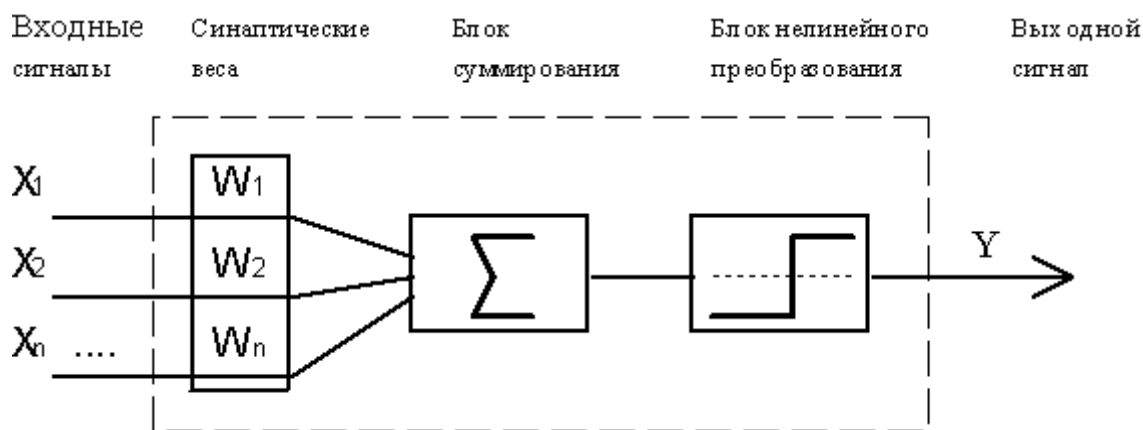


Рис. 8. Формальный нейрон Маккалоха и Питтса

В блоке суммирования происходит накопление общего входного сигнала (обычно обозначаемого символом *net*), равного взвешенной сумме входов:

$$net = \sum_i w_i x_{ij} \quad (8)$$

Геометрическая интерпретация формального нейрона – гиперплоскость. Отклик нейрона далее описывается по принципу "все или ничего", т. е.

переменная подвергается нелинейному пороговому преобразованию, при котором выход (состояние активации нейрона)  $Y$  устанавливается равным единице, если  $net > p$ , и  $Y=0$  в обратном случае.

$$Y = \begin{cases} 1 & net > p, \\ 0 & net < p. \end{cases} \quad (9)$$

Искусственная нейронная сеть представляет собой систему соединённых и взаимодействующих между собой формальных нейронов. Каждый нейрон подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам. И, тем не менее, будучи соединёнными в достаточно большую сеть с управляемым взаимодействием, такие локально простые нейроны вместе способны выполнять довольно сложные задачи.

Приведем пример нейронной сети прямого распространения, которая была использована в работе «Применение нейронных сетей к задачам оценки вероятности дефолта по потребительским кредитам» [90].

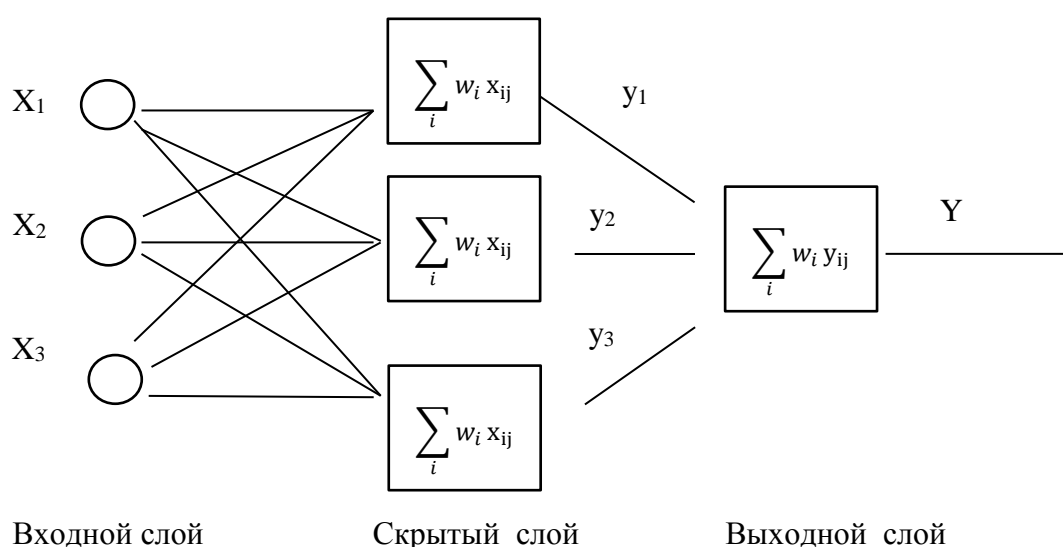


Рис. 9. Архитектура нейронной сети прямого распространения

Нейронные сети не программируются в привычном смысле этого слова, они обучаются. Возможность обучения — одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение заключается в нахождении коэффициентов связей между нейронами. В процессе обучения нейронная сеть способна выявлять сложные зависимости между



входными данными и выходными, а также выполнять обобщение. Это значит, что в случае успешного обучения сеть сможет вернуть верный результат на основании данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искаженных данных [18,175,176].

В книге Нейронные сети полный курс [18] показано, что нейронная сеть способна описать нелинейную функцию любой сложности, в связи с чем нейронные сети обладают наиболее высоким качеством классификации заемщиков по сравнению с остальными методами. Так, Bank of New York опробовав метод нейронных вычислений и применив его для оценки 100 тыс. банковских счетов, выявил свыше 90,0 % потенциальных неплательщиков [17].

Основной проблемой применения нейронных сетей при анализе заемщиков является то, что нейронная сеть является «черным ящиком» в связи с чем затруднительно определить влияние параметров заемщика на вероятность дефолта по кредиту [67,70]. Невозможность четко выделить зависимость параметров заемщика и кредитного риска препятствует активному внедрению нейронных сетей, во многом из-за того, что не представляется возможным сформировать кредитную стратегию организации. Заметим, что в США нейронные сети практически не применяются для оценки кредитных рисков при выдаче кредита, в связи законодательным требованием предоставить заемщику причины, повлекшие за собой отказ от выдачи кредита [17].

## **Глава 2. Применение комитетных конструкций к классификации заемщиков**

### **2.1. Классификация заемщиков физических лиц. Формулировка задачи классификации заемщиков методом комитета**

Определение кредитоспособности заемщика является неотъемлемой частью работы банка по определению возможности выдачи кредита. Под анализом кредитоспособности заемщика понимается оценка банком возможности и целесообразности предоставления заемщику кредитов, определения вероятности их своевременного возврата в соответствии с кредитным договором [14,17].

В большинстве кредитных организаций, ориентированных на потребительское кредитование, действуют автоматизированные системы принятия решений, так называемые «кредитные фабрики». При принятии решения по выдаче кредитные организации рассматривают такие параметры заемщика, как платежеспособность, возраст, место работы, должность, наличие имущества, и иные показатели в случае необходимости.

Рассмотрим задачу принятия решения по выдаче кредита на кредитном комитете простым большинством голосов. От кредитного инспектора на кредитный комитет поступает кредитное заявление, содержащее информацию о сумме и сроке кредита, поле, возрасте, доходах заемщика, имуществе заемщика и другие параметры заемщика. Каждым членом комитета, исходя из собственного опыта, анализируются полученные данные и высказывается мнение о возможности выдачи кредита или отказ в выдаче кредита. Решение принимается простым большинством голосов [62,63,64,65,89].

В рамках данной работы комитет, в котором принятие решения осуществляется простым большинством голосов, будет называться комитетом большинства.

Сформулируем задачу принятия решения о выдаче кредита комитетом большинства из  $q$  равноправных членов. Пусть имеются  $m$  заемщиков каждому из которых может быть принято решение выдать кредит или отказать в выдаче кредита. Договоримся нумеровать членов комиссии индексом  $t$ , а заемщиков индексом  $j$ . Допустим, что  $j$ -му заемщику соответствует множество параметров  $x_j$ , а у каждого  $p$ -го эксперта имеется множество предпочтений в выдаче кредита в зависимости от параметров заемщика, обозначенные  $w^t$ . В случае если параметры заемщика соответствуют предпочтениям эксперта, то есть  $x_j \in w^t$  данный эксперт голосует за выдачу кредита.

Членом комитета будем называть некоторое подпространство в пространстве признаков  $w^t$ , комитетом называется некоторое множество членов комитета  $Q = (w^1, w^2, \dots, w^q)$ . Через  $|\{t | x_j \in w^t\}|$  обозначим процедуру голосования, то есть количество членов комитета, голосующих за выдачу кредита.

Задача классификации заемщиков методом комитетов заключается в нахождении такого множества членов комитета, применение которого позволяет разделить заемщиков на 2 класса.

При принятии комитетом большинства из  $q$  членов решения по  $j$  заемщику, в случае если  $|\{t | x_j \in w^t\}| > \frac{q}{2}$  то принимается решение выдать кредит, в случае если  $|\{t | x_j \in w^t\}| < \frac{q}{2}$  принимается решение отказать в выдаче кредита [42].

Так же существуют различные виды комитетов, такие как комитет единогласия в котором решение за или против должно приниматься единогласным решением комитета, то есть при принятии комитетом единогласия из  $q$  членов решения по  $j$  заемщику, в случае если  $|\{t | x_j \in w^t\}| = q$  то принимается решение выдать кредит, в случае если  $|\{t | x_j \in w^t\}| < q$  принимается решение отказать в выдаче кредита.

Более общим видом комитета является  $p$ -комитет в котором изначально не устанавливается число членов, которые должны проголосовать за или против, а определяются в процессе построения комитетного решения. При принятии  $p$ -

комитетом из  $q$  членов решения по  $j$  заемщику, в случае если  $|\{t|x_j \in w^t\}| \geq p$  то принимается решение выдать кредит, в случае если  $|\{t|x_j \in w^t\}| < p$  принимается решение отказать в выдаче кредита, где  $p = [1..q]$  [16,46].

## **2.2. Метод комитетов. Формулировка задачи принятия решения методом комитетов**

Выбор и диагностика — два фундаментальных понятия в моделировании широкого круга явлений — физико-технических, экономических, социальных и других. Эти понятия являются базисными в теории принятия решений.

Комитетные конструкции можно рассматривать и как некоторый класс обобщений понятия решения на случай несовместных систем уравнений, неравенств и включений, и как средство распараллеливания при решении задач выбора, диагностики и прогнозирования. Как обобщение понятия решения задачи комитетные конструкции представляют собой наборы элементов, обладающие некоторыми (но, как правило, не всеми) свойствами решения: это вид размытых решений. Например, комитет системы ограничений — это такой набор элементов, что каждому ограничению удовлетворяет более половины элементов набора [8,9,10,11,12,13,16,21,22].

Метод комитетов определяет одно из направлений исследований задач оптимизации и классификации. Он связан с построением конструкций, обобщающих понятие решения системы уравнений и неравенств, позволяющих наряду с разрешимыми задачами анализировать задачи с противоречивыми условиями. В чистом виде одна из простейших комитетных конструкций появилась в 1965 году как способ построения некоторой обучающейся нейронной сети [101]. После этого данная тема широко исследовалась, но чисто математическая теория с полными и строгими доказательствами развивалась в основном в Институте математики и механики УрО РАН (г. Екатеринбург) [43,44,45,47,48,49,50,70,82,86,87]. Однако стоит заметить, что изучение несовместных задач, обладающих ясным практическим смыслом, имеет

достаточно давнюю историю. Так, еще Лежандр, Гаусс, а позднее Лаплас при разработке метода наименьших квадратов имели дело с переопределенными, а значит, как правило, противоречивыми, системами линейных уравнений [42].

Метод комитетов использует построение составных решающих правил диагностики и выбора, имеющих вид голосования базисных решающих правил. При этом, в частности, используется правило принятия решений по большинству голосов. В теории принятия решений правило большинства впервые, видимо, исследовал Кондорсе. Он, в частности, показал нетранзитивность коллективных предпочтений, устанавливаемых по правилу большинства. Вопросы корректности процедур голосования исследовались в дальнейшем многими авторами [29,30,31,32,34,35].

Противоречивые системы уравнений и неравенств в моделях оптимизации и классификации возникают закономерно, они не являются результатами случайных ошибок или логически некорректных рассуждений. Это просто реальные постановки задач, которые являются только исходным пунктом для построения конструкций, определенным строго оговоренным способом разрешающих противоречивые системы ограничений [33,83,84,85].

Приведем условный геометрический пример классификации заемщиков методом комитетов [58]. Пусть в банке существует унифицированный кредитный продукт с фиксированной стоимостью и сроком (условно кредит в 1 млн. руб. на 3 года), тогда заемщиков можно характеризовать 2 параметрами = заявленный месячный доход и возраст.

Представим множество заемщиков на плоскости:

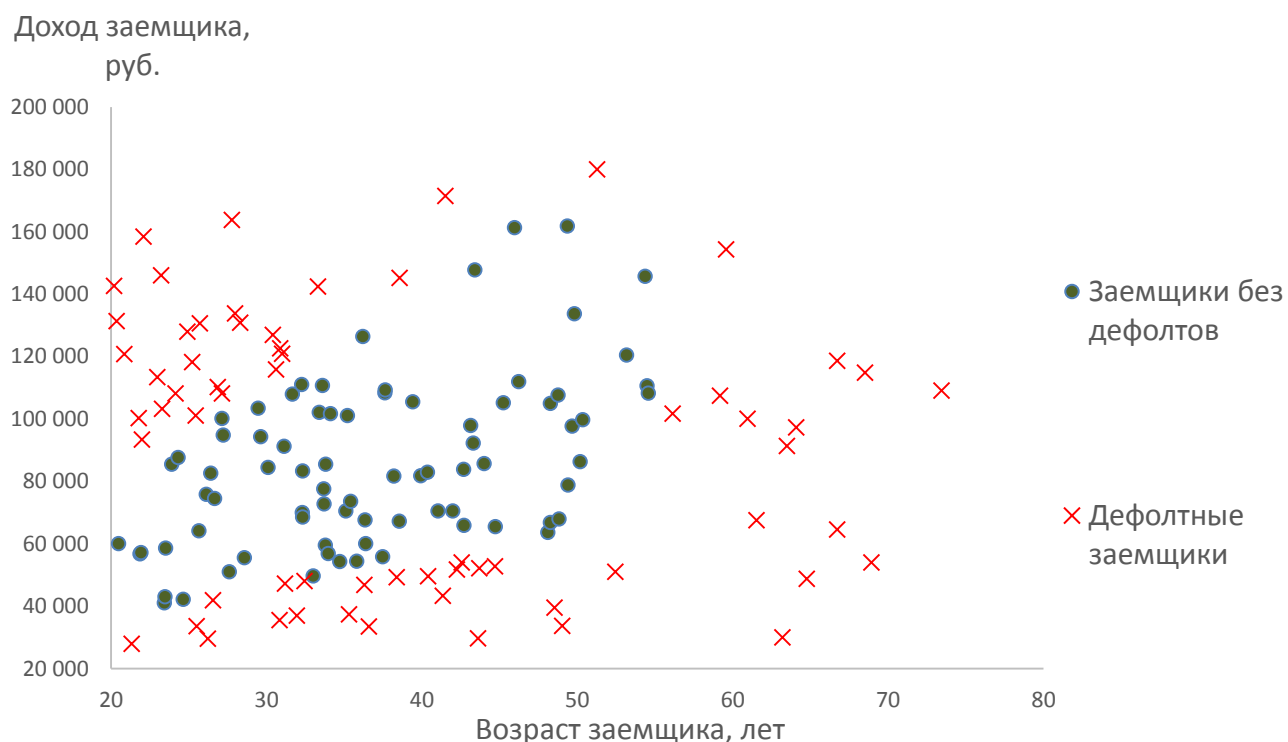


Рис 10. Геометрический пример

На рисунке 10 по оси Y указан доход заемщика, по оси X указан возраст заемщика, заемщики делятся на 2 класса – заемщики со случаем дефолта через 12 месяцев (красные крестики), заемщики без случаев дефолта (синие точки).

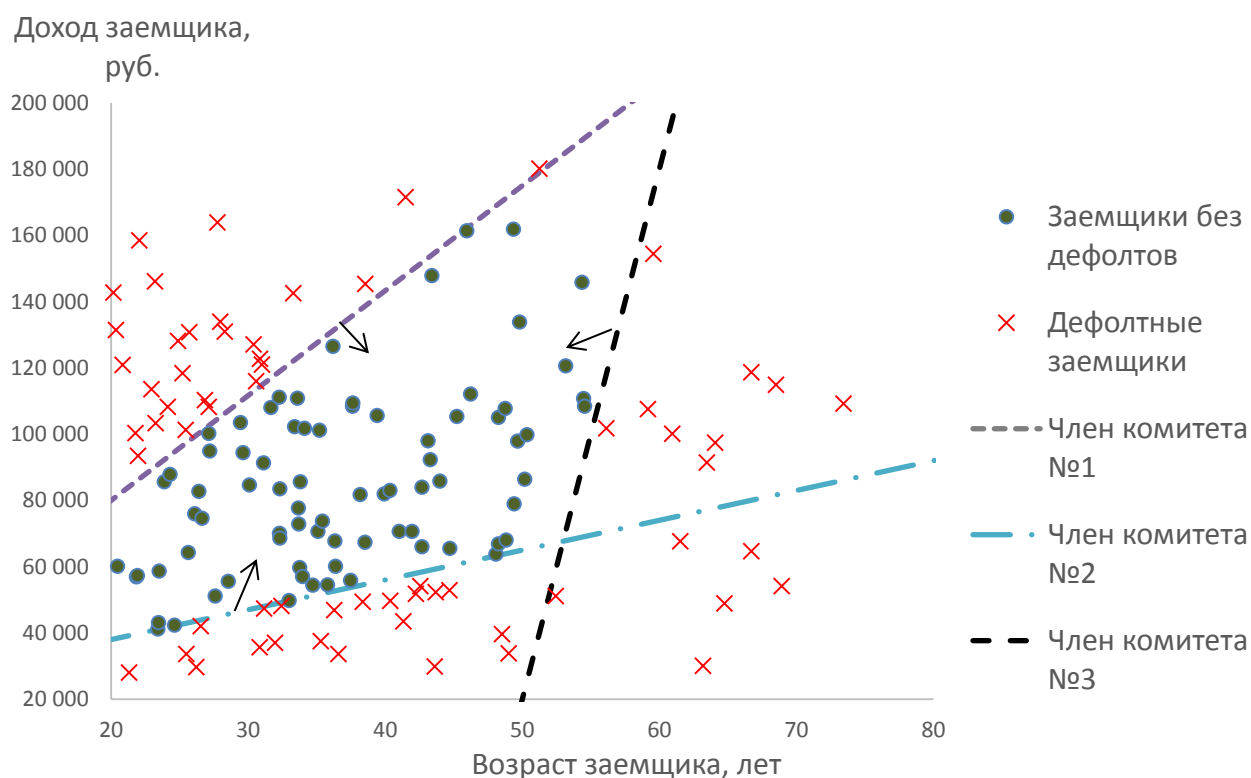


Рис 11. Классификация заемщиков методом комитетов

Как видно из рисунка 11 комитет из 3 членов позволяет точно классифицировать заемщиков. Стрелочками на графике указано направление «голосования» члена комитета за отсутствие дефолта. Член комитета №1 «отсекает» заемщиков, у которых указанные в заявке доходы не соответствуют возрасту, член комитета №2 «отсекает» заемщиков с низкими доходами, член комитета №3 «отсекает» заемщиков в возрасте.

Заметим, что для практических задач не всегда целесообразно использовать комитетные решения, имеющие на обучающие выборке 100% качество классификации, на что указывается в работе Вл. Д. Мазурова [16]. Определение оптимального числа членов комитета определяется в рамках процессов верификации и валидации моделей.

Наиболее подробно метод построения комитета, предложенный Вл. Д. Мазурова описан в работах [16,34,35,36,37,38,40,41,43,44,45].

### **2.3. Построение комитета через решение задачи математического программирования**

Задачу построения комитета можно свести к задаче частично-целочисленного программирования. В данной работе для построения комитетов применяется пакет IBM ILOG CPLEX Optimization Studio.

В рамках данной работы к построению комитета будет применена теория математического программирования, в частности задача построения комитета будет сведена к задаче линейного частично-целочисленного программирования. Отметим, что ранее задачи частично-целочисленного программирования не применялись для построения комитетных конструкций. Далее будет приведено последовательное сведение задачи построения комитета к задаче частично-целочисленного программирования. Основным элементом метода комитета являются члены комитета — линейные дискриминанты. Рассмотрим задачу построение линейного дискриминанта через решение задачи математического программирования.

### 2.3.1 Решение задачи построения линейного дискриминанта через решение задачи математического программирования

Математически задачу с построения линейной гиперплоскости сформулировать следующим образом. Пусть заданы множество  $M \subset R^n$  и класс функций  $F \subset \{R^n \rightarrow R\}$ . Известно, что  $M = K_1 \cup K_2$ , причем множества  $K_1$  и  $K_2$  заданы своими конечными подмножествами  $A \subset K_1$ ,  $B \subset K_2$ . Задачей построения дискриминанта называется задача нахождения функции  $f \in F$  такой, что:

$$\begin{cases} f(a) > 0, \forall a \in A, \\ f(b) < 0, \forall b \in B. \end{cases} \quad (10)$$

Найдем такую функцию  $f$ , что  $K_1 = \{x \in M \mid f(x) > 0\}$  и  $K_2 = \{x \in M \mid f(x) \leq 0\}$ .

Пусть  $x_j \in K_1$ , где  $j \in J$ ,  $x_i \in K_2$ , где  $i \in I$ ,  $J$ -число элементов первого множества,  $I$ - число элементов второго множества.

Поскольку  $M \subset R^n$ , где  $n$  является размерностью пространства (число параметров, характеризующих заемщика), то договоримся обозначать  $x \in M$ , как  $x_{z,j} \in K_1$ ,  $x_{z,i} \in K_2$ , где  $z \in Z$ ,  $x_{z,j}$  – параметры конкретного заемщика.

Через  $Z$  обозначается множество  $\{1, 2, \dots, n\}$ .

Тогда задача дискриминации множества на подмножества заключается в построении такой гиперплоскости, для которой выполняется следующая система линейных неравенств:

$$\begin{cases} \sum_z w_z * x_{z,j} - c > 0, \forall j \in J, \\ \sum_z w_z * x_{z,i} - c < 0, \forall i \in I. \end{cases} \quad (11)$$

Где  $w_1, w_2, \dots, w_n$  являются коэффициентами в линейном неравенстве, а  $c$ , соответственно, является свободным членом линейного неравенства.

Приведем графический пример классификации линейным дискриминантом – гиперплоскостью, в случае, если рассматриваемая система ограничений (11) является совместной.



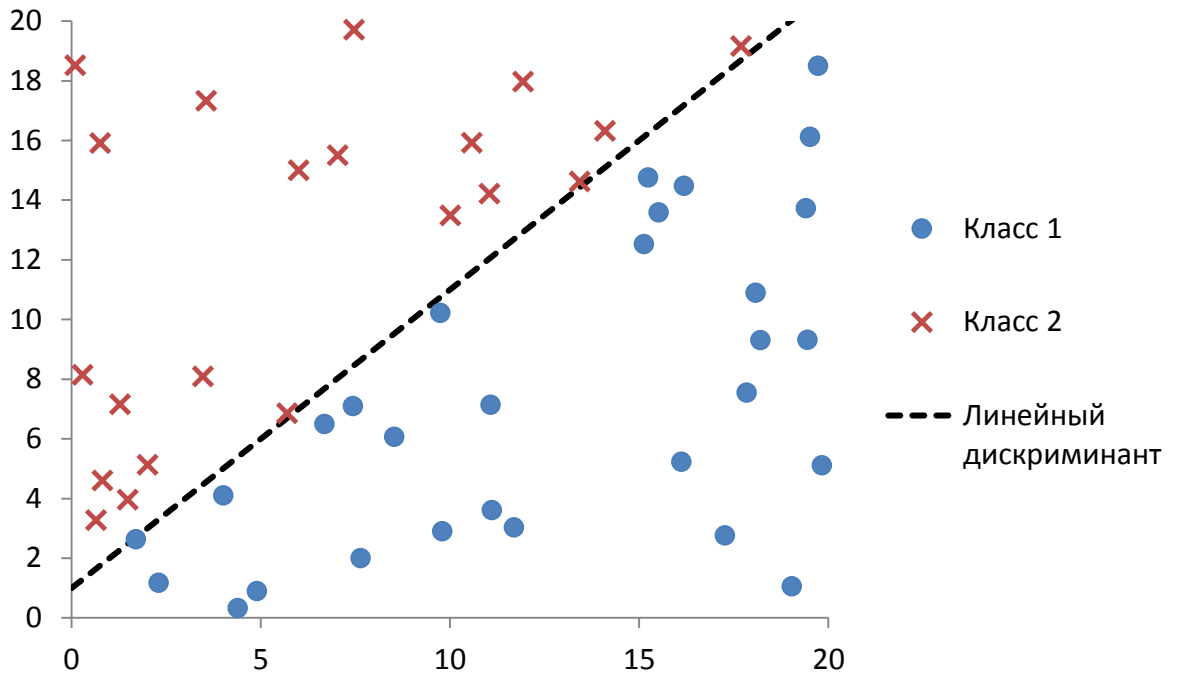


Рис. 12. Линейный дискриминант для совместной системы ограничений

Как уже было сказано выше, данную гиперплоскость можно построить только в случае, если система неравенств (11) является совместной при каких либо  $w_z, c$ .

В случае если система является несовместной для построения разделяющей гиперплоскости необходимо ввести множества "невязок"  $V = (v_1, v_2, \dots, v_j)$  и  $V' = (v'_1, v'_2, \dots, v'_i)$ . Отметим, что множества "невязок"  $V$  и  $V'$  строго положительны.

Тогда задача построения гиперплоскости сводится к задаче линейного программирования следующего вида:

$$\begin{cases} \sum_z w_z * x_{z,j} - c - v_j > 0, \forall j \in J, \\ \sum_z w_z * x_{z,i} - c + v'_i < 0, \forall i \in I \end{cases} \quad (12)$$

$$\min \sum_j v_j + \sum_i v'_i$$

Приведем графический пример классификации линейным дискриминантом — гиперплоскостью, в случае, если рассматриваемая система ограничений является несовместной.

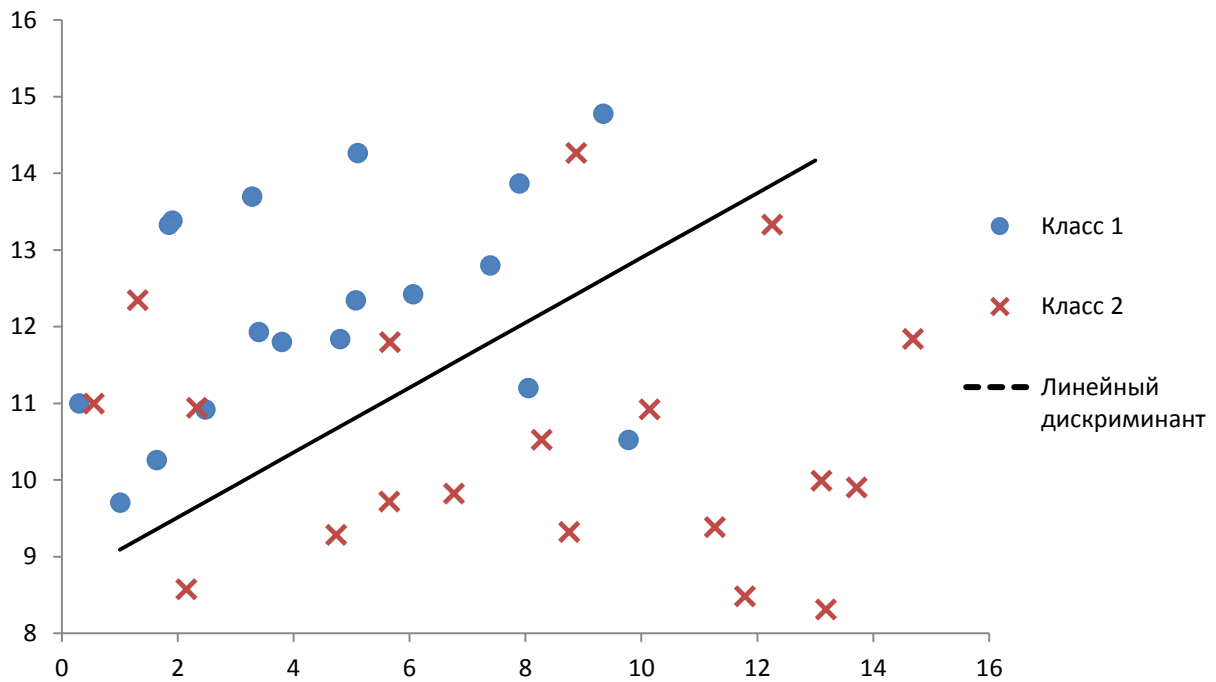


Рис. 13. Линейный дискриминант для несовместной системы ограничений

Как видно из рис. 13 в случае, если система является несовместной, не представляется возможным отделить множество точек одного класса от другого за счет использования линейного дискриминанта.

Поскольку в случае несовместности системы (11) множество неразделимо одной гиперплоскостью, необходимо использовать метод комитетов.

### 2.3.2 Построение комитета большинства, комитета единогласия и $p$ -комитета через решение задачи математического программирования.

Математически метод комитета большинства можно сформулировать следующим образом. Пусть заданы множество  $M \subset R^n$  и класс функций  $F \subset \{R^n \rightarrow R\}$ . Известно, что  $M = K_1 \cup K_2$ , причем множества  $K_1$  и  $K_2$  заданы своими конечными подмножествами  $A \subset K_1$ ,  $B \subset K_2$ . Задачей построения комитета большинства называется задача нахождения набора функций  $f = (f_1, \dots, f_q), f \in F$  такого, что :

$$\begin{cases} |\{t | f_t(a) > 0\}| > \frac{q}{2}, \quad \forall a \in A, \\ |\{t | f_t(b) < 0\}| > \frac{q}{2}, \quad \forall b \in B. \end{cases} \quad (13)$$

где  $q$  – число членов комитета.

Пусть  $x_j \in K_1$ , где  $j \in J$ ,  $x_i \in K_2$ , где  $i \in I$ ,  $J$ -число элементов первого множества,  $I$ - число элементов второго множества.

Поскольку  $M \subset R^n$ , где  $n$  является размерностью пространства (число параметров, характеризующих заемщика), то договоримся обозначать  $x \in M$ , как  $x_{z,j} \in K_1$ ,  $x_{z,i} \in K_2$ , где  $z \in Z$ .

$t$ -ым членом комитета будем называть последовательность  $(w_1^t, w_2^t, \dots, w_Z^t)$ . Введем множество "невязок"  $V = (v_1^t, v_2^t, \dots, v_j^t)$  и  $V' = (v_1'^t, v_2'^t, \dots, v_i'^t)$ .  $t$  – индекс, указывающий на принадлежность к члену комитета.

Для построения комитета большинства для всех  $j, i$  должно быть справедливо неравенство:

$$\sum_t v_j^t + \sum_t v_i'^t < \frac{q}{2}, t \in (1, 2, \dots, q) \quad (14)$$

$$V, V' \in \{0, 1\}.$$

То есть, сумма "невязок" для каждого элемента подмножеств  $K_1, K_2$  должно быть меньше половины числа членов комитета.

Задача построения комитета большинства для разделения множества на подмножества сводится к решению следующей системы уравнений:

$$\begin{cases} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\ \sum_t v_j^t + \sum_t v_i'^t < \frac{q}{2} \end{cases} \quad (15)$$

где,  $L$  – некоторое большое число.

Приведем графический пример классификации комитетом большинства.

Пусть у нас имеются следующее множество точек, разделенное на 2 класса на плоскости.

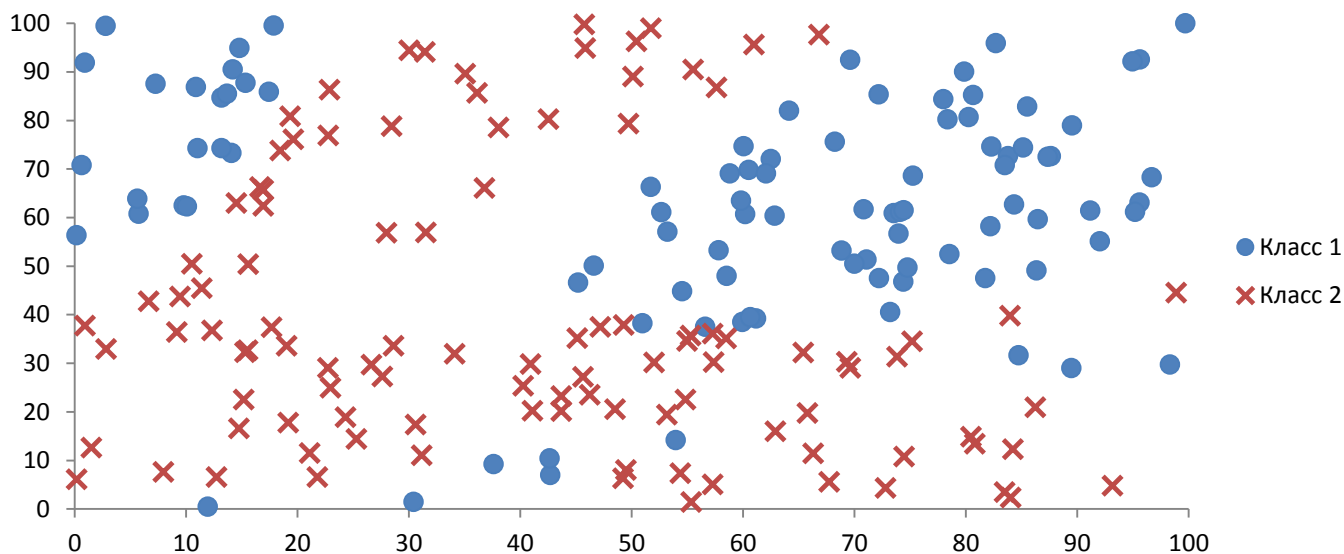


Рис. 14. Множество точек, разделенное на 2 класса

Для того чтобы разделить данные множества построим комитет из 5 членов.

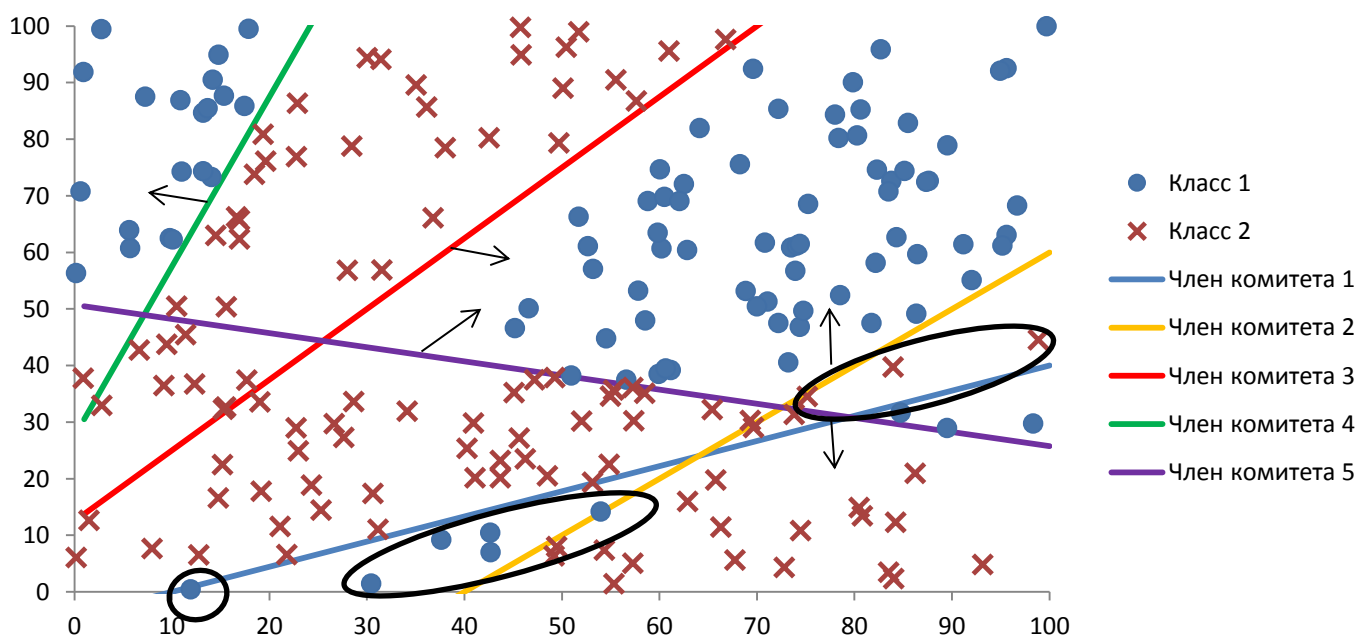


Рис. 15. Комитет большинства, разделяющий множество на 2 класса.

Стрелочками на рис. 15 показаны направления голосования членов комитета за принадлежность наблюдения 1 классу.

Заметим, что если  $\nexists x_{z,j} \in K1, x_{z,j} \in K2$ , где  $j \in J, z \in Z$  и  $\nexists x_{z,i} \in K1, x_{z,i} \in K2$ , где  $i \in I, z \in Z$ , то есть ни одно наблюдение не принадлежит одновременно двум множествам, то возможно построение комитета большинства из  $q$ -членов.

Заметим, что для практических задач не всегда целесообразно использовать комитетные решения, имеющие на обучающей выборке 100% качество классификации, поскольку в исходных данных часто имеются ошибки и шумы, в связи с чем не смотря на 100% классификацию на обучающей выборке, качество решения на тестовой выборке значительно снижается, на что указывается в работе Вл. Д. Мазурова [16].

Как видно из рис. 4 построенный комитет большинства позволил со равным 100% качеством разделить множество на 2 класса. Кружочками обведены те наблюдения, которые скорее можно объяснить ошибками в данных, шумом и недостаточностью переменных.

В связи с тем, что построение комитета со 100% качеством классификации на обучающей выборке не целесообразно перейдем к построению комитета большинства, который будет иметь максимальное качество классификации (минимальное число ошибок классификации) на обучающей выборке.

Введем дополнительные множества невязок  $Y = (y_1, y_2, \dots, y_j)$  и  $Y' = (y'_1, y'_2, \dots, y'_i)$ ,  $Y, Y' \in \{0,1\}$ .

Запишем задачу построения комитета большинства как задачу частично-целочисленного программирования:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i^t < 0, \forall i \in I \\ \sum_t v_j^t < \frac{q}{2} + L * y_j \\ \sum_t v_i^t < \frac{q}{2} + L * y'_j \\ \min \sum_j y_j + \sum_i y'_i \end{array} \right. \quad (16)$$

где,  $L$  – некоторое большое число.

Для множества точек, приведенного на рис. 4 построим комитет из 3 членов.

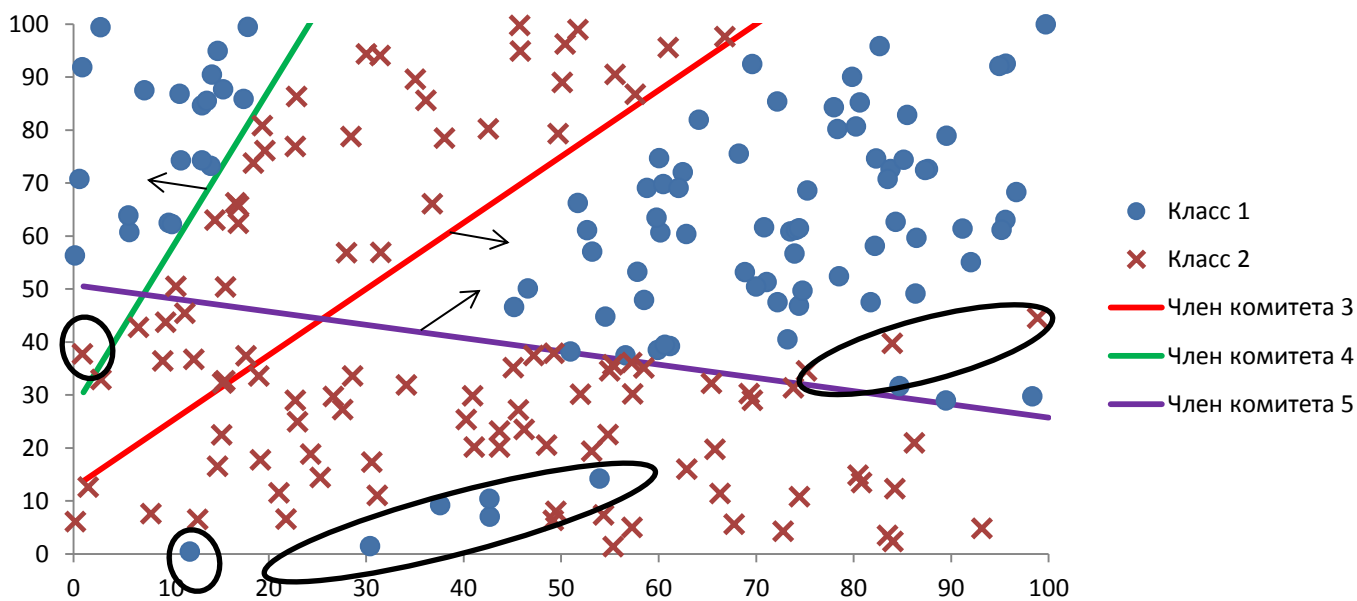


Рис. 16. Комитет большинства из 3 членов

Стрелочками на рис. 16 показаны направления голосования членов комитета за принадлежность наблюдения 1 классу. Кружочками обведены ошибки классификации.

Как видно из рис. 16 построенный комитет из 3 членов не позволяет с точностью 100% разделить множество на 2 класса, но при этом точность данного комитета при практическом применении может быть выше, чем точность комитета, который разделяет со 100% точностью.

Далее приведем сведение задачи построения комитета единогласия через решение задачи частично-целочисленного линейного программирования.

Напомним, что комитетом единогласия называется такой комитет в котором за или против принадлежности точек к 2 классам голосуют все члены комитета.

Математически метод комитета единогласия можно сформулировать следующим образом. Пусть заданы множество  $M \subset R^n$  и класс функций  $F \subset \{R^n \rightarrow R\}$ . Известно, что  $M = K_1 \cup K_2$ , причем множества  $K_1$  и  $K_2$  заданы своими конечными подмножествами  $A \subset K_1$ ,  $B \subset K_2$ . Задачей построения комитета единогласия называется задача нахождения набора функций  $f = (f_1, \dots, f_q), f \in F$  такого, что :

$$\begin{cases} |\{t|f_t(a) > 0\}| = q, \quad \forall a \in A, \\ |\{t|f_t(b) < 0\}| > 0, \quad \forall b \in B. \end{cases} \quad (17.1)$$

или

$$\begin{cases} |\{t|f_t(a) > 0\}| > 0, \quad \forall a \in A, \\ |\{t|f_t(b) < 0\}| = q, \quad \forall b \in B. \end{cases} \quad (17.2)$$

где  $q$  – число членов комитета. В системе неравенств 17.1 приведен пример, когда все члены комитета единогласно голосуют за принадлежность наблюдения к 1 классу, а в системе неравенств 17.2 приведен пример, когда все члены голосуют за принадлежность наблюдения к классу 2.

Пусть  $x_j \in K_1$ , где  $j \in J$ ,  $x_i \in K_2$ , где  $i \in I$ ,  $J$ -число элементов первого множества,  $I$ - число элементов второго множества.

Поскольку  $M \subset R^n$ , где  $n$  является размерностью пространства (число параметров, характеризующих заемщика), то договоримся обозначать  $x \in M$ , как  $x_{z,j} \in K_1$ ,  $x_{z,i} \in K_2$ , где  $z \in Z$ .

Через  $Z$  обозначается множество  $\{1, 2, \dots, n\}$ .

$t$ -ым членом комитета будем называть последовательность  $(w_1^t, w_2^t, \dots, w_z^t)$ . Введем множество "невязок"  $V = (v_1^t, v_2^t, \dots, v_j^t)$  и  $V' = (v_1'^t, v_2'^t, \dots, v_i'^t)$ .  $t$  – индекс, указывающий на принадлежность к члену комитета.

Для построения комитета единогласия в случае, если все члены комитета должны проголосовать за принадлежность к подмножеству  $K_1$ , для всех  $i$  должно быть справедливо неравенство:

$$\sum_t v_i'^t < q, t \in (1, 2, \dots, q) \quad (18.1.)$$

$$V' \in \{0, 1\}.$$

Для построения комитета единогласия в случае, если все члены комитета должны проголосовать за принадлежность к подмножеству  $K_2$ , для всех  $j$  должно быть справедливо неравенство:

$$\sum_t v_j^t < q, t \in (1, 2, \dots, q) \quad (18.2.)$$

$$V \in \{0, 1\}.$$

Задача построения комитета единогласия для разделения множества на подмножества, если все члены комитета должны проголосовать за принадлежность к подмножеству  $K_I$ , сводится к решению следующей системы уравнений:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\ \sum_t v_i'^t < q \end{array} \right. \quad (19.1)$$

Задача построения комитета единогласия для разделения множества на подмножества, если все члены комитета должны проголосовать за принадлежность к подмножеству  $K_2$ , сводится к решению следующей системы уравнений:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c < 0, \forall i \in I. \\ \sum_t v_j^t = 0 \end{array} \right. \quad (19.2)$$

Пусть у нас имеются следующее множество точек, разделенное на 2 класса на плоскости.



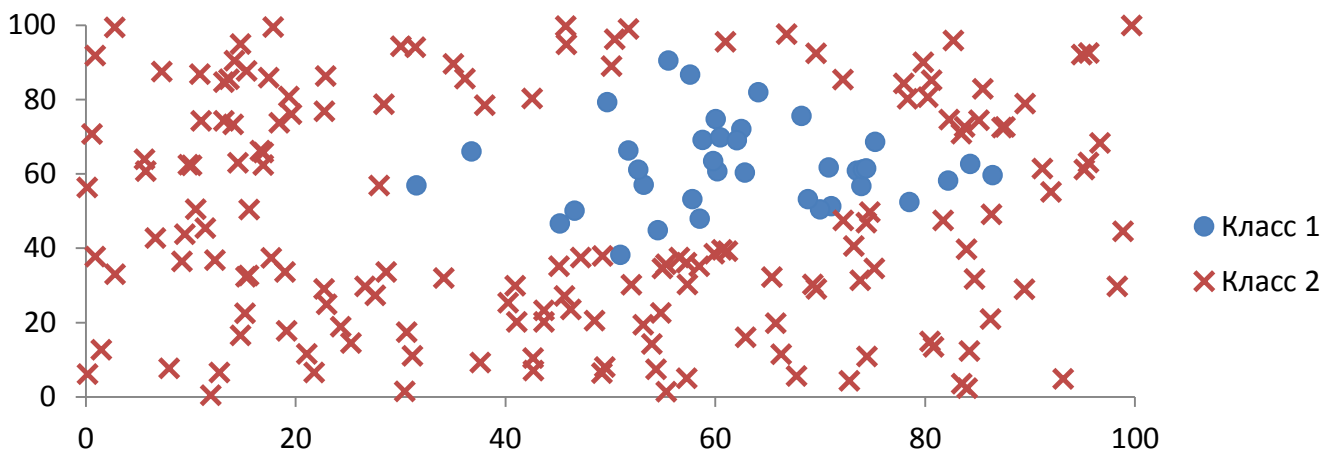


Рис. 17. Множество точек, разделенное на 2 класса

Для того чтобы разделить данные множества построим комитет единогласия из 4 членов.

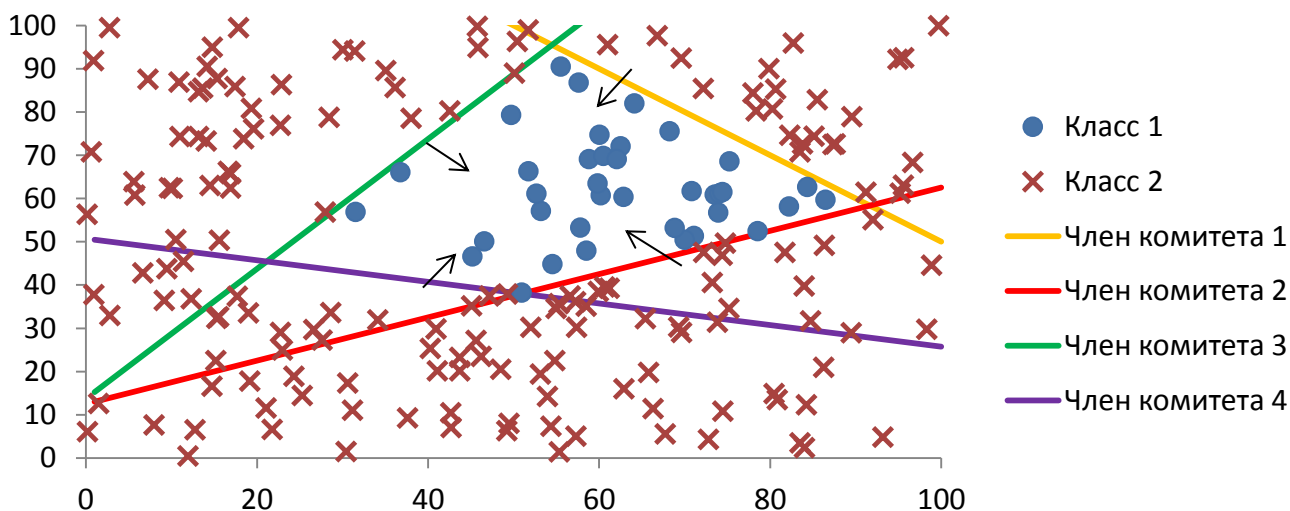


Рис. 18. Разделение множества комитетом единогласия из 4 членов.

Стрелочками на рис. 18 показаны направления голосования членов комитета за принадлежность наблюдения 1 классу. Как видно из рис. 8 данное множество эффективно разделяется комитетом единогласия из 4 членов, тогда как для разделения данного множества комитетом большинства потребовалось бы значительно большее число членов.

Заметим, что комитет большинства всегда можно построить, если  $\nexists x_{z,j} \in K1, x_{z,j} \in K2$ , где  $j \in J, z \in Z$  и  $\nexists x_{z,i} \in K1, x_{z,i} \in K2$ , где  $i \in I, z \in Z$ , то есть ни одно наблюдение не принадлежит одновременно двум множествам, тогда как очевидно, что комитет единогласия позволяющий разделить множество на 2

класса о 100% только в том случае, если в выпуклой оболочке одного из классов отсутствуют точки другого класса.

В связи с тем, что комитет единогласия, позволяющий 100% классификацию можно построить не в каждом случае, перейдем к построению комитета единогласия, который будет иметь максимальное качество классификации (минимальное число ошибок классификации) на обучающей выборке.

Введем дополнительные множества невязок  $Y = (y_1, y_2, \dots, y_j)$  и  $Y' = (y'_1, y'_2, \dots, y'_i)$ ,  $Y, Y' \in \{0,1\}$ .

Запишем задачу построения комитета единогласия, если все члены комитета должны проголосовать за принадлежность к подмножеству  $K_I$ , как задачу частично-целочисленного линейного программирования:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\ \sum_t v_j^t \leq L * y_j \\ \sum_t v_i'^t < q + L * y_j' \\ \min \sum_j y_j + \sum_i y_i' \end{array} \right. \quad (20.1)$$

Запишем задачу построения комитета единогласия, если все члены комитета должны проголосовать за принадлежность к подмножеству  $K_2$ , как задачу частично-целочисленного линейного программирования:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i^t < 0, \forall i \in I \\ \sum_t v_j^t < q + L * y_j \\ \sum_t v_i^t \leq L * y_j' \\ \min \sum_j y_j + \sum_i y_i' \end{array} \right. \quad (20.2)$$

Приведем геометрический пример построения комитета единогласия с качеством классификации менее 100%

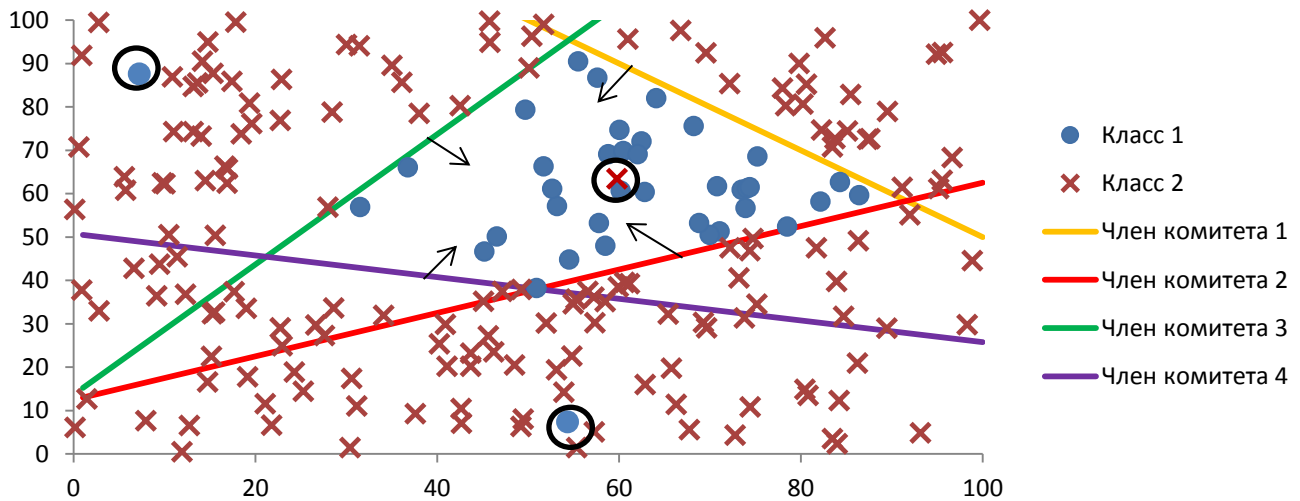


Рис. 19. Разделение множества комитетом единогласия из 4 членов с качеством классификации менее 100%.

Стрелочками на рис. 19 показаны направления голосования членов комитета за принадлежность наблюдения 1 классу. Кругочками обведены ошибки классификации.

Выше были рассмотрены комитет большинства, в котором решения принимаются большинством голосов, и комитет единогласия, в котором решения принимаются единогласно. На начальном этапе не представляется возможным определить какой именно из комитетов целесообразнее применять для классификации множеств. Заметим так же, что существуют случаи когда решения целесообразнее применять не большинством голосов или единогласно, а

допустим в соотношении 6 из 7, то есть для определения принадлежности наблюдения к классу 1 достаточно положительного решения 6 из 7 членов, в связи с чем перейдем к описанию р-комитета.

Математически р-комитет можно сформулировать следующим образом. Пусть заданы множество  $M \subset R^n$  и класс функций  $F \subset \{R^n \rightarrow R\}$ . Известно, что  $M = K_1 \cup K_2$ , причем множества  $K_1$  и  $K_2$  заданы своими конечными подмножествами  $A \subset K_1$ ,  $B \subset K_2$ . Задачей построения р-комитета называется задача нахождения набора функций  $f = (f_1, \dots, f_q), f \in F$  такого, что:

$$\begin{cases} |\{t|f_t(a) > 0\}| > p, & \forall a \in A, \\ |\{t|f_t(b) < 0\}| \geq q - p, & \forall b \in B. \end{cases} \quad (21)$$

где  $p = [0..q - 1]$ .

Тогда, в случае если  $p = \frac{q}{2}$  возникает система неравенств соответствующая комитету большинства:

$$\begin{cases} |\{t|f_t(a) > 0\}| > \frac{q}{2}, & \forall a \in A, \\ |\{t|f_t(b) < 0\}| \geq \frac{q}{2}, & \forall b \in B. \end{cases} \quad (22.1)$$

Тогда, в случае если  $p = 0$  (11.2) или  $p = q - 1$  (11.3) системы неравенств соответствуют комитету единогласия:

$$\begin{cases} |\{t|f_t(a) > 0\}| > 0, & \forall a \in A, \\ |\{t|f_t(b) < 0\}| \geq q, & \forall b \in B. \end{cases} \quad (22.2)$$

$$\begin{cases} |\{t|f_t(a) > 0\}| > q - 1, & \forall a \in A, \\ |\{t|f_t(b) < 0\}| \geq 1, & \forall b \in B. \end{cases} \quad (22.3)$$

Пусть  $x_j \in K_1$ , где  $j \in J$ ,  $x_i \in K_2$ , где  $i \in I$ ,  $J$ -число элементов первого множества,  $I$ - число элементов второго множества.

Поскольку  $M \subset R^n$ , где  $n$  является размерностью пространства (число параметров, характеризующих заемщика), то договоримся обозначать  $x \in M$ ,

как  $x_{z,j} \in K_1, x_{z,i} \in K_2$ , где  $z \in Z$ .

Через  $Z$  обозначается множество  $\{1, 2, \dots, n\}$ .

$t$ -ым членом комитета будем называть последовательность  $(w_1^t, w_2^t, \dots, w_z^t)$ . Введем множество "невязок"  $V = (v_1^t, v_2^t, \dots, v_j^t)$  и  $V' = (v_1'^t, v_2'^t, \dots, v_i'^t)$ .  $t$  – индекс, указывающий на принадлежность к члену комитета.

Для построения  $p$ -комитета для всех  $j, i$  должно быть справедливо неравенство:

$$\begin{cases} \sum_t v_j^t \leq p \\ \sum_t v_i'^t < q - p \end{cases} \quad (23)$$

где  $p = [0..q - 1]$ .

Задача построения  $p$ -комитета сводится к решению следующей системы неравенств.

$$\begin{cases} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\ \sum_t v_j^t \leq p \\ \sum_t v_i'^t < q - p \\ p \geq 0 \\ p \leq q - 1 \end{cases} \quad (24)$$

Приведем графический пример классификации.

Пусть у нас имеется следующее множество точек, разделенных на 2 класса в двумерном пространстве.

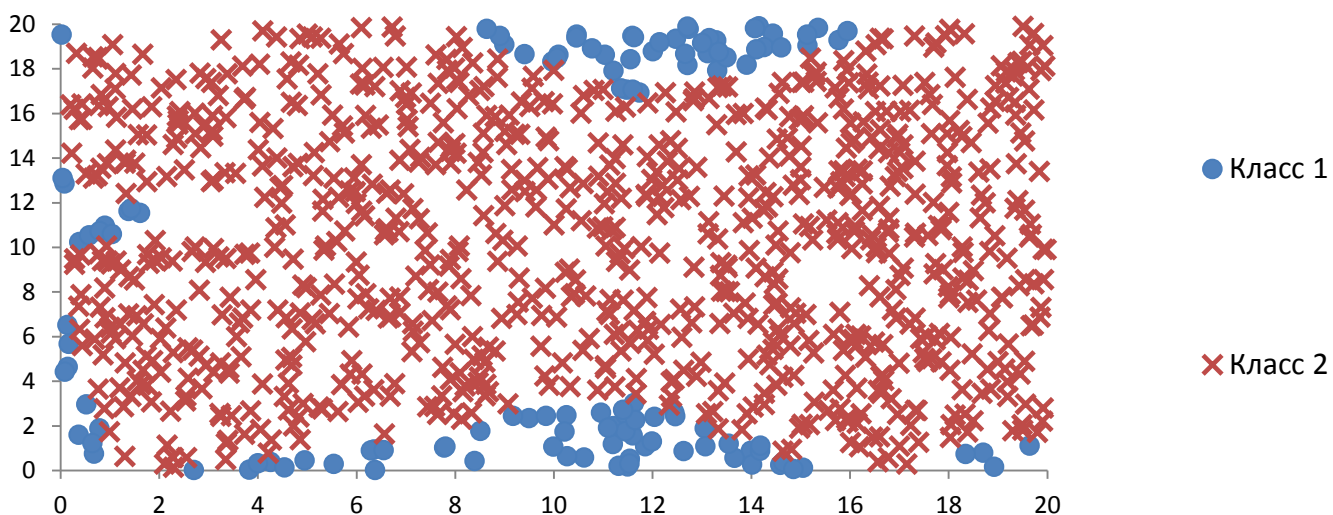


Рис. 20. Множество точек, разделенное на 2 класса

Попробуем разделить множество  $p$ -комитетом из 7 членов, в котором  $p$  принимает значения отличные от  $q/2$ , 0 и  $q-1$ .

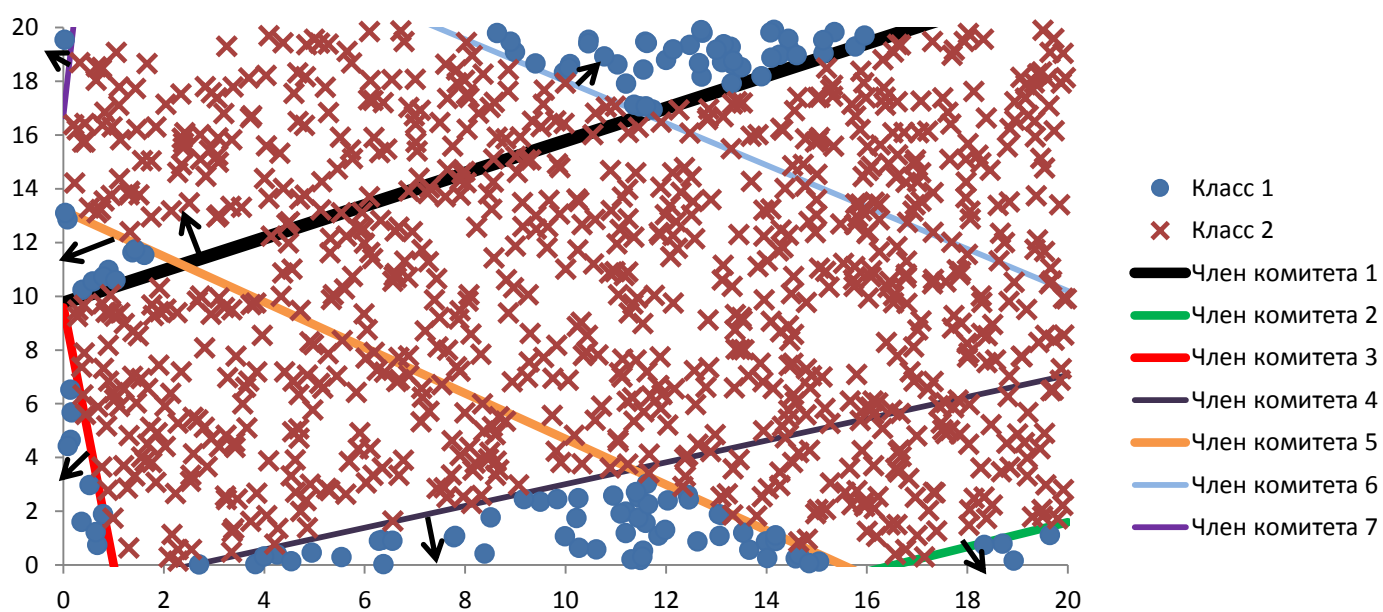


Рис. 21. Разделение множества  $p$ -комитетом из 5 членов ( $p=5$ ).

Стрелочками на рис. 21 показаны направления голосования членов комитета за принадлежность наблюдения 1 классу.

Как видно из рис. 21 для того, чтобы точка относилась к 1 классу достаточно, чтобы за нее проголосовало 2 или более членов комитета (то есть  $p=5$ ).

Как уже было указано ранее, при решении практических задач не всегда целесообразно, чтобы точность на обучающей выборке была 100%, в связи с чем

перейдем к решению задачи построения  $p$ -комитета, который будет иметь максимальное качество классификации (минимальное число ошибок классификации) не всегда равное 100% на обучающей выборке.

Введем дополнительные множества невязок  $Y = (y_1, y_2, \dots, y_j)$  и  $Y' = (y'_1, y'_2, \dots, y'_i)$ ,  $Y, Y' \in \{0,1\}$ .

Запишем задачу построения  $p$ -комитета как задачу частично-целочисленного программирования:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\ \sum_t v_j^t \leq p + L * y_j \\ \sum_t v_i'^t < q - p + L * y'_j \\ p \geq 0 \\ p \leq q - 1 \\ \min \sum_j y_j + \sum_i y'_i \end{array} \right. \quad (25)$$

Для решения задачи классификации заемщиков в данной работе будет применяться  $p$ -комитет, в связи с тем, что изначально неизвестно какой именно комитет (единогласия, большинства или с соотношением 5 из 7)

В рассматриваемом  $p$ -комитете целевая функция определена как функция ошибок комитета при классификации и стоит задача минимизации числа ошибок комитета.

Далее в работе будут приведен пример иной целевой функции, в рамках которой, задачей будет являться минимизация числа ошибок комитета при одновременной минимизации числа ошибок каждого из членов комитета.

## 2.3 Выбор целевой функции. Сравнение результатов комитетных решений в зависимости от выбранной целевой функции

При сведении задачи построения  $p$ -комитета к задаче частично-целочисленного программирования целевая функция была определена как функция минимизации числа ошибок классификации, в данной части будет рассмотрена функция минимизации числа ошибок классификации при одновременной минимизации числа ошибок членов комитета.

Заметим, что при решении практических задач классификации, в многомерном пространстве могут быть построены несколько  $p$ -комитетов с одинаковым числом членов, которые имеют одинаковое качество классификации, в связи с чем возникает задача определения наиболее оптимального  $p$ -комитета. В рамках данной работы наиболее оптимальным будет признаваться комитет, в котором при равном качестве классификации будет наблюдаться меньшее число ошибок классификации у членов комитета.

Напомним, что изначально задача построения  $p$ -комитета сводилась к задаче частично-целочисленного программирования (16), при этом целевая функция определена как:

$$\min \sum_j y_j + \sum_i y'_i \quad (26)$$

Сформулируем целевую функцию, как функцию нахождения минимального числа ошибок классификации, при минимальном числе ошибок членов комитета, при этом «штраф» за ошибку классификации комитетом должен превосходить «штраф» за ошибку классификации членом комитета. Тогда задача частично целочисленного программирования сводится к следующей задаче:



$$\begin{cases}
\sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\
\sum_z w_z^t * x_{z,i} - c + L * v_i^t < 0, \forall i \in I \\
\sum_t v_j^t \leq p + L * y_j \\
\sum_t v_i^t < q - p + L * y'_j \\
p \geq 0 \\
p \leq q - 1
\end{cases} \quad (27)$$

$$\min L' * (\sum_j y_j + \sum_i y'_i) + \sum_t \sum_j v_j^t + \sum_t \sum_i v_i^t$$

где,  $L'$  - некоторое большое число, превосходящее число наблюдений на обучающей выборке, умноженное на число членов комитета. Очевидно, что при большом  $L'$  «штраф» за ошибочную классификацию комитетом будет превосходить штраф за ошибочную классификацию членами комитета. Следовательно, при построении комитета из 2 комитетов имеющих одинаковое качество классификации будет выбран комитет, который имеет меньшее число ошибок членов комитета.

Заметим, что задача частично-целочисленного программирования подходит для решения задач классификации в случае, если число наблюдений одного класса близко к числу наблюдений другого класса. В рассматриваемой нами задаче классификации заемщиков число наблюдений относящихся к заемщикам в дефолте составляет менее 2% от общего числа заемщиков. Следовательно, если построение комитета будет проводиться через решение задачи линейного частично-целочисленного программирования (26), то минимальное общее число ошибок может быть достигнуто в том случае, если комитет корректно определит все наблюдения без случаев дефолта (98% наблюдений), но при этом некорректно определить большую часть наблюдений со случаями дефолта (2% наблюдений). Данное комитентов решение не будет иметь практического применения.

Для решения задачи построения комитета, в случае если число наблюдений одного класса значительно меньше числа наблюдений другого класса, необходимо взвесить ошибки классификации, исходя из числа наблюдений каждого класса.

Введем дополнительные константы  $R_1$  и  $R_2$ , равные числу наблюдений 1 и 2 классов соответственно. Тогда задача построения  $p$ - комитета сводится к задаче частично целочисленного программирования следующего вида:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\ \sum_t v_j^t \leq p + L * y_j \\ \sum_t v_i'^t < q - p + L * y'_j \\ p \geq 0 \\ p \leq q - 1 \end{array} \right. \quad (28)$$

$$\min L' * (\sum_j y_j * R_1 + \sum_i y'_i * R_2) + \sum_t \sum_j v_j^t * R_1 + \sum_t \sum_i v_i'^t * R_2$$

В рамках данной работы для построения  $p$ -комитета, будет решаться вышеприведенная задача частично-целочисленного программирования (28).

Для подтверждения целесообразности использования целевой функции минимизирующей одновременно ошибки классификации комитета и членов комитета (28), а не целевой функции минимизирующей только ошибки комитетного решения, проведем сравнение качества решений при использовании различных целевых функций.

Имеется обучающая выборка, содержащая информацию о 3000 заемщиков без случаев дефолта и 60 заемщиков без случаев дефолта.

Было построено 2 комитета: первый - с использованием целевой функции, приведенной в формуле 15, и второй - с использованием целевой функции, приведенной в формуле 16. Естественно в обе целевые функции введено

взвешивание ошибок классификации, исходя из числа наблюдений каждого класса.

В таблице 2 приведены результаты комитетного решения на обучающей выборке.

Таблица 2. Качества комитетного решения на обучающей выборке

Голосование комитета	Наблюдения со случаями дефолта	Наблюдения без случаев дефолта	Доля заемщиков в дефолте
«За» дефолт	44	1 285	3,4%
«Против» дефолта	16	1 715	0,9%
Итого	60	3 000	2,0%

Качества первого и второго комитета на обучающей выборке совпали.

В таблице 3 приведены результаты качества комитетного решения на тестовой выборке для первого комитета, в котором целевая функция – только минимизация числа ошибок классификации комитета.

Таблица 3. Качество комитетного решения (целевая функция 26) на тестовой выборке

Голосование комитета	Наблюдения со случаями дефолта	Наблюдения без случаев дефолта	Доля заемщиков в дефолте
«За» дефолт	281	10 439	2,7%
«Против» дефолта	477	27 462	1,7%
Итого	758	37 901	2,0%

Как видно из таблицы 2 доля заемщиков в дефолте в случае, если комитет голосует «за» дефолт, на тестовой выборке снизилась с 3,4% до 2,7%, что показывает снижение качества комитетного решения.

В таблице 4 приведены результаты качества комитетного решения первого комитета, в котором целевая функция – минимизация числа ошибок классификации комитета при одновременной минимизации числа ошибок членов комитета.

Таблица 4. Качество комитетного решения (целевая функция 28) на тестовой выборке

Голосование комитета	Наблюдения со случаями дефолта	Наблюдения без случаев дефолта	Доля заемщиков в дефолте
«За» дефолт	300	9 215	3,3%
«Против» дефолта	458	28 686	1,6%
Итого	758	37 901	2,0%

Как видно из таблицы 4 доля заемщиков в дефолте в случае, если комитет голосует «за» дефолт, на тестовой выборке снизилась с 3,4% до 3,3%.

Исходя из таблиц 3 и 4, отмечается меньшее снижение качества классификации в случае применения целевой функции, минимизирующей одновременно и ошибки классификации комитета и ошибки классификации каждого из членов комитета.

### 2.3.1 Комитет с неравными весами членов

В представленных ранее комитетных конструкциях предполагается, что веса каждого из членов комитета одинаковы, однако можно предположить, что возможно увеличение качества классификации за счет взвешивания членов комитета. Рассмотрим задачу построения  $r$ -комитета, приведенного в формуле 11, веса членов которого будут иметь разный вес при принятии комитетного решения.

Задача построения  $r$ -комитета с разными весами членов комитета состоит в нахождении набора функций  $f = (f_1, \dots, f_q), f \in F$  и вектора весов  $\varphi = (\varphi_1, \dots, \varphi_q)$  таких, что:

$$\left\{ \begin{array}{l} \sum_{t=1}^q |\{t|f_t(a) > 0\}| * \varphi_t > p, \quad \forall a \in A, \\ \sum_{t=1}^q |\{t|f_t(b) < 0\}| * \varphi_t \geq 1 - p, \quad \forall b \in B, \\ \sum_{t=1}^q \varphi_t = 1 \end{array} \right. \quad (29)$$

где  $p = [0..1], \varphi_t = [0..1]$ .

Напомним, что ранее задача построения  $p$ -комитета сводилась к задаче линейного частично-целочисленного программирования следующего вида:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\ \sum_t v_j^t \leq p + L * y_j, \quad \forall j \in J \\ \sum_t v_i'^t < q - p + L * y_i', \quad \forall i \in I \\ p \geq 0 \\ p \leq q - 1 \\ \min \sum_j y_j + \sum_i y_i' \end{array} \right. \quad (30)$$

Задача построения комитета с неравными веса членов комитета сводится к следующей оптимизационной задаче.

$$\begin{cases}
\sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\
\sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\
\sum_t v_j^t * \varphi^t \leq p + L * y_j, \quad \forall j \in J \\
\sum_t v_i'^t * \varphi^t < 1 - p + L * y'_j, \forall i \in I \\
p \geq 0, \\
p \leq q - 1, \\
\sum_t \varphi^t = 1.
\end{cases} \quad (31)$$

$$\min \sum_j y_j + \sum_i y'_i$$

На первый взгляд решение задачи построения р-комитета с разными весами членов комитета не может быть сведено к задаче линейного частично-целочисленного программирования, постольку для решения указанной задачи необходимо осуществлять умножение двух переменных. Для сведения задачи 19 к задаче линейного частично-целочисленного программирования заменим выражения  $v_j^t * \varphi^t$  и  $v_i'^t * \varphi^t$  на переменные  $\varphi_j^t$  и  $\varphi_i'^t$  такие, что:

$$\begin{aligned}
v_j^t = 0 &\Rightarrow \varphi_j^t = 0, v_j^t = 1 \Rightarrow \varphi_j^t = \varphi^t, \\
v_i'^t = 0 &\Rightarrow \varphi_i'^t = 0, v_i'^t = 1 \Rightarrow \varphi_i'^t = \varphi^t
\end{aligned} \quad (32)$$

Для выполнений условий, указанных выше, должна выполняться следующая система неравенств.

$$\begin{cases}
\varphi_j^t \geq \varphi^t + v_j^t - 1, \\
\varphi_j^t \leq v_j^t, \\
\varphi_j^t \leq \varphi^t, \\
\varphi_j^t \geq 0.
\end{cases} \quad (33)$$

Аналогичная система неравенств должна выполняться и для  $\varphi_i'^t$ .

Проведем проверку системы неравенств, пусть некоторое  $v_j^t = 0$ , тогда система неравенств 21 выглядит следующим образом:

$$\begin{cases} \varphi_j^t \geq \varphi^t - 1, \\ \varphi_j^t \leq 0, \\ \varphi_j^t \leq \varphi^t, \\ \varphi_j^t \geq 0. \end{cases} \quad (34)$$

Очевидно, что в случае, если  $v_j^t$  принимает значение 0, то  $\varphi_j^t$  принимает значение 0.

Пусть  $v_j^t = 1$ , тогда система неравенств 21 выглядит следующим образом:

$$\begin{cases} \varphi_j^t \geq \varphi^t, \\ \varphi_j^t \leq 1, \\ \varphi_j^t \leq \varphi^t, \\ \varphi_j^t \geq 0. \end{cases} \quad (35)$$

Очевидно, что в случае, если  $v_j^t$  принимает значение 1, то  $\varphi_j^t$  принимает значение  $\varphi^t$ .

Поскольку произведения  $v_j^t * \varphi^t$  и  $v_i'^t * \varphi^t$  были заменены на непрерывные переменные  $\varphi_j^t$  и  $\varphi_i'^t$ , то запишем задачу построения комитета с разными весами членов комитета как задачу линейного частично-целочисленного программирования.

$$\left\{ \begin{array}{l}
\sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\
\sum_z w_z^t * x_{z,i} - c + L * v_i'^t < 0, \forall i \in I \\
\sum_t \varphi_j^t \leq p + L * y_j, \quad \forall j \in J \\
\sum_t \varphi_i'^t < 1 - p + L * y'_j, \forall i \in I \\
\varphi_j^t \geq \varphi^t, \\
\varphi_j^t \leq 1, \\
\varphi_j^t \leq \varphi^t, \\
\varphi_j^t \geq 0, \\
\varphi_i'^t \geq \varphi^t, \\
\varphi_i'^t \leq 1, \\
\varphi_i'^t \leq \varphi^t, \\
\varphi_i^t \geq 0, \\
p \geq 0, \\
p \leq q - 1, \\
\sum_t \varphi^t = 1.
\end{array} \right. \quad (36)$$

$$\min \sum_j y_j + \sum_i y'_i$$

Заметим, что целевую функцию также можно определить, как:

$$\min L' * (\sum_j y_j + \sum_i y'_i) + \sum_t \sum_j \varphi_j^t + \sum_t \sum_i \varphi_i'^t$$

Где  $L'$  некоторое большое число. Ранее в разделе 2.3. была показана практическая применимость введения в целевую функцию ошибок членов комитета.

Приведем графический пример, на котором построение комитета с неравными весами возможно с меньшим числом членов, чем при построении комитета с равными весами.



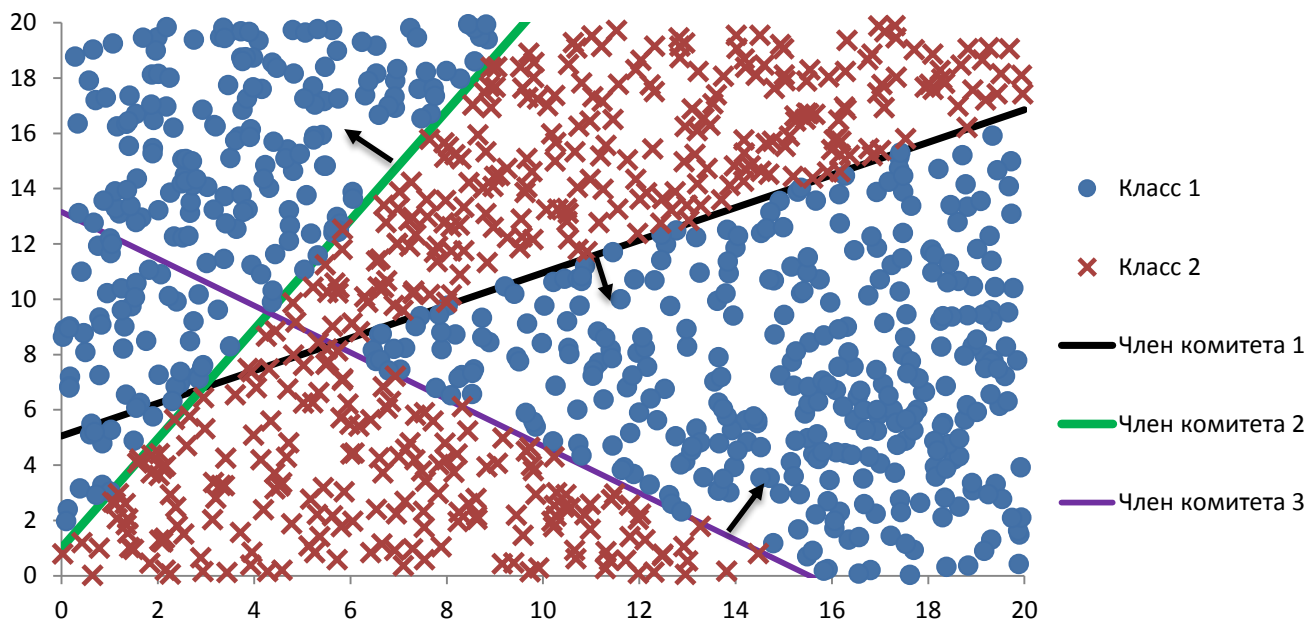


Рис.22. Комитет с неравными весами членов.

Стрелочками на рисунке 22 указано направление голосования члена комитета за класс 1. Вес члена комитета 2 определен как 0,5, веса остальных членов комитета равны 0,25. Комитетное правило определено следующим образом, если сумма весов членов комитета проголосовавших за 1 класс больше или равна 0,5, то присваивается 1 класс, иначе присваивается класс 2. Для решения данной задачи  $p$ -комитетом с равными весами членов потребовалось бы построить комитет из 4 членов, в котором 2 члена комитета бы практически совпадали.

Заметим, что практическая применимость комитета с неравными весами членов для решаемой в данной работе задачи практически отсутствует в связи со значительным увеличением сложности решения задачи.

## 2.4 Оценка сложности построения комитета. Нахождение начального решения задачи частично-целочисленного программирования

Для построения комитета необходимо решить задачу линейного частично-целочисленного программирования. В данной части работы мы постараемся оценить сложность построения комитета, в том числе будут приведены данные о времени затраченном для построения комитета при решении практических задач.

Общее число целочисленных переменных при решении задачи построения  $p$ -комитета по формуле 17 равняется  $(q + 1) * n + 1$ , где  $q$ -число членов

комитета,  $n$ - общее число наблюдений обоих классов. Число непрерывных переменных равняется  $(W + 1) * q$ , где  $W$  – число параметров наблюдений.

Так, например, если требуется построить комитет из 7 членов на выборке, содержащей 100 наблюдений, при этом у каждого наблюдения 5 параметров, то необходимо решить задачу линейного частично-целочисленного программирования с 801 целочисленной переменной и 42 непрерывными переменными.

Заметим, что большинство задач целочисленного программирования (ЦП) относится к классу NP-трудных, и для их решения не существует эффективных полиномиальных алгоритмов [53], то есть сложность решения задачи построения комитета будет возрастать с увеличением числа членов комитета и увеличением числа наблюдений, при этом заметим, что увеличение числа членов увеличивает число целочисленных членов на  $n$ , в связи с чем при решении практических задач нами не использовались комитеты, содержащие более 11 членов. Решение задач построения комитета в рамках данной работы осуществляется в пакете IBM ILOG CPLEX, в котором для решения задач частично-целочисленного программирования используется метод ветвей и границ, а так же многочисленные эвристические методы.

В таблице далее приведены данные о времени, которое было затрачено, для построения комитетов при классификации заемщиков, и точности комитетного решения. Нами решалась задача построения комитета содержащего 4500 заемщиков без случаев дефолта и 150 заемщиков со случаями дефолта, число параметров описывающих заемщиков равнялось 6.

Таблица 5. Время, затраченное на построение комитета.

Число членов комитета	Точность классификации на обучающей выборке, %	Число целочисленных переменных	Затраченное время, сек.
3	53,3%	18601	1530
5	55,0%	27901	3272

7	64,2%	37201	7655
9	66,5%	46501	18354
11	68,9%	55801	34563

Как видно из таблицы 5 точность комитетного решения на обучающей выборке возрастает с увеличением числа членов комитета, при этом с увеличением числа членов комитета значительно возрастает время, которое необходимо затратить для построения комитета. Так для построения комитета из 3 членов было затрачено 1530 сек. (4.25 часов), тогда как для построения комитета из 11 членов потребовалось 34563 сек. (101,92 часов), то есть увеличение числа членов в 3,67 раза приводит к увеличению времени необходимого для построения комитета в 24,0 раза.

Как было отмечено выше, для решения задачи частично-целочисленного программирования в пакете IBM ILOG CPLEX применяется метод ветвей и границ. При применении данного метода решения целочисленных задач важным для времени решения задачи является выбор начального решения задачи, при котором при целочисленных значениях переменных выполняются все ограничения задачи. При решении задачи классификации заемщиков пакет IBM ILOG CPLEX определяет начальное решение, при котором точность классификации одного множества равняется 100%, а другого множества 0%, то есть строится комитет единогласия, в который включаются все наблюдения обучающей выборки. Заметим, что пакет IBM ILOG CPLEX позволяет пользователю самостоятельно определить начальное решение задачи. Выбор начального решения с более высоким качеством классификации позволяет значительно снизить время которое необходимо затратить для решения задачи частично-целочисленного программирования.

### *2.5.1 Применение эволюционного алгоритма для нахождения начального решения задачи частично-целочисленного программирования*

Для нахождения начального решения задачи частично-целочисленного программирования можно применить эвристические алгоритмы оптимизации, такие как эволюционные (генетические) алгоритмы [102,119].

Эволюционные алгоритм - это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных естественному отбору в природе.

Идея применения генетических алгоритмов основана на эволюционной теории, согласно которой каждый биологический вид с течением времени совершенствует свои качества так, чтобы эффективней справляться с важнейшими задачами выживания, самозащиты, размножения и т.д. Таким путем возникла защитная окраска у многих рыб и насекомых, панцирь у черепахи, яд у скорпиона и многие другие полезные приспособления.

С помощью эволюции природа постоянно оптимизирует все живое, находя неординарные решение. Неясно, за счет чего происходит этот прогресс, однако ему можно дать научное пояснение, базируясь всего на двух биологических механизмах - естественного отбора и генетического наследования.

Отметим, что при применении эволюционных алгоритмов возникает проблема того, что эволюционные алгоритмы имеют тенденцию сходиться к локальному оптимуму или даже к спорным точкам, вместо глобального оптимума для данной задачи, в связи с чем применение данных алгоритмов к решению практических задач представляется спорным. [102]

Для нахождения начального решения задачи частично-целочисленного программирования в рамках данной работы применяется эволюционный алгоритм, основанный на «мутациях». Под мутациями понимается случайное изменение «генов» - параметров комитетов. На начальном этапе генерируется 1000 случайных комитетов, параметры членов комитета определяются исходя из равномерного распределения от -1 до 1, далее осуществляется отбор 5 комитетов

(0,5% от общего числа сгенерированных комитетов), имеющих минимальное значение целевой функции, определенной в задаче частично-целочисленного программирования 17. На следующем этапе происходит мутация параметров комитетов, то есть к каждому из параметров комитетов, отобранных на предыдущем этапе, добавляется случайное значение равномерной функции распределения от -0,1 до 0,1, для каждого комитета данная операция выполняется 200 раз, далее осуществляется отбор 0,5% комитетов, имеющих минимальное значение целевой функции. Операция мутации параметров и отбора комитетов осуществляется 1000 раз. Нахождение начального решения для комитета из 7 членов при 6 параметрах, описывающих заемщика, занимает около 30 минут.

Рассмотрим задачу построения комитета единогласия за счет применения эволюционного алгоритма. Пусть у нас имеются следующее множество точек, разделенное на 2 класса на плоскости.

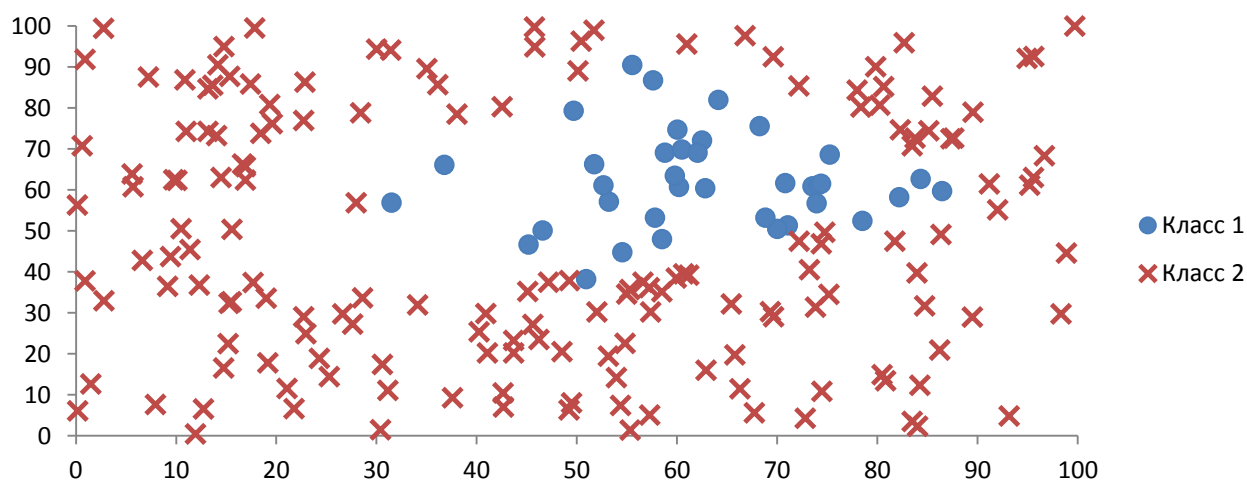


Рис. 23. Множество точек, разделенное на 2 класса

На начальном этапе сгенерируем 1000 случайных комитетов из 4 членов. На рисунке 23 приведен комитет, имеющий минимальное значение целевой функции.

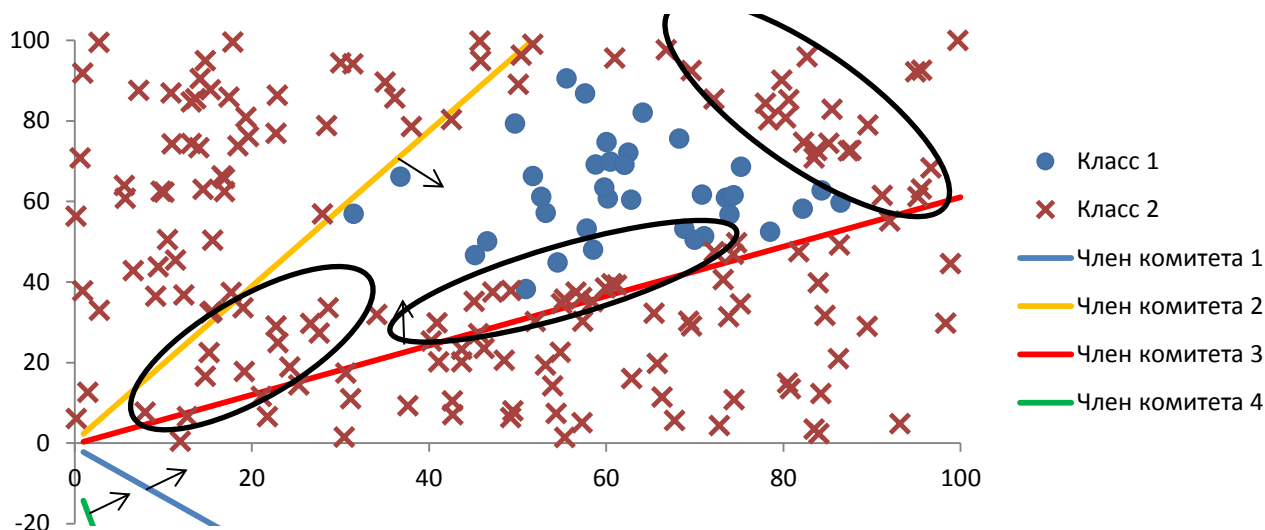


Рис. 24. Разделение множества комитетом из 4 членов, сгенерированным на начальном этапе.

Стрелочками на рисунке 24 отмечены направления голосования комитетов за принадлежность к классу 1. Как видно из рисунка случайно сгенерированный комитет имеет 100% классификацию класса 1, ошибки классификации наблюдений класса 2 обведены кругами. Значение целевой функции составляет 249 195.

На рисунке 14 приведен комитет единогласия из 4 членов, полученный через 200 операций мутаций и параметров комитетов.

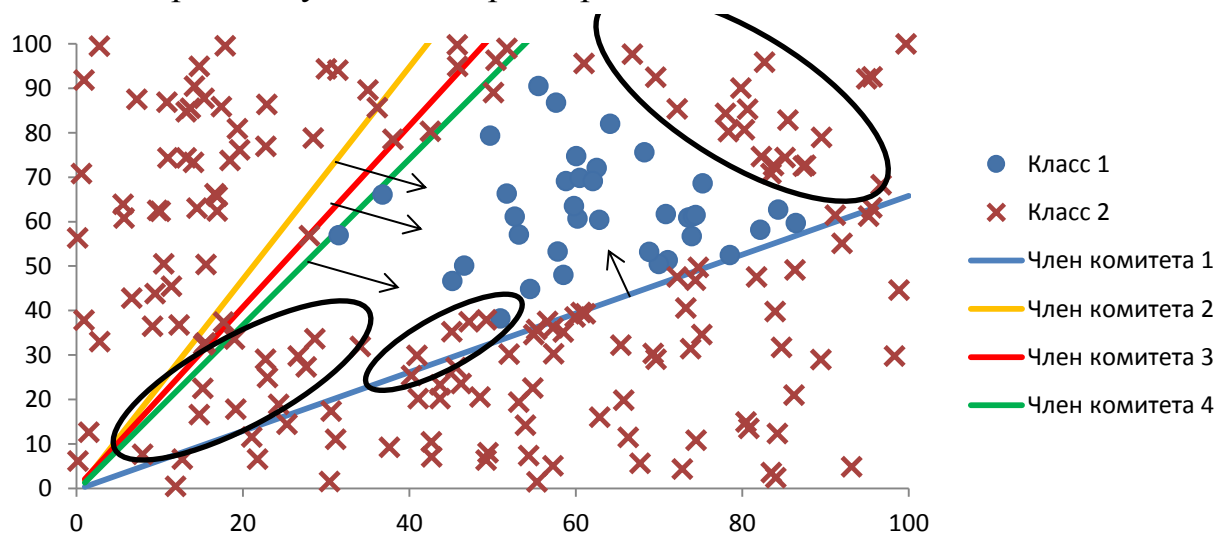


Рис. 25. Разделение множества комитетом из 4 членов, сгенерированным после 200 операций мутаций и отбора.

Как видно из рисунка 25, после 200 операций мутаций и отбора качество классификации комитета из 4 членов возросло, значение целевой функции при

этом составило 181 670, напомним, что на начальном этапе значение целевой функции составило 249 195.

Далее рассмотрим комитет из 4 членов, полученный после 1000 операций мутаций и отбора.

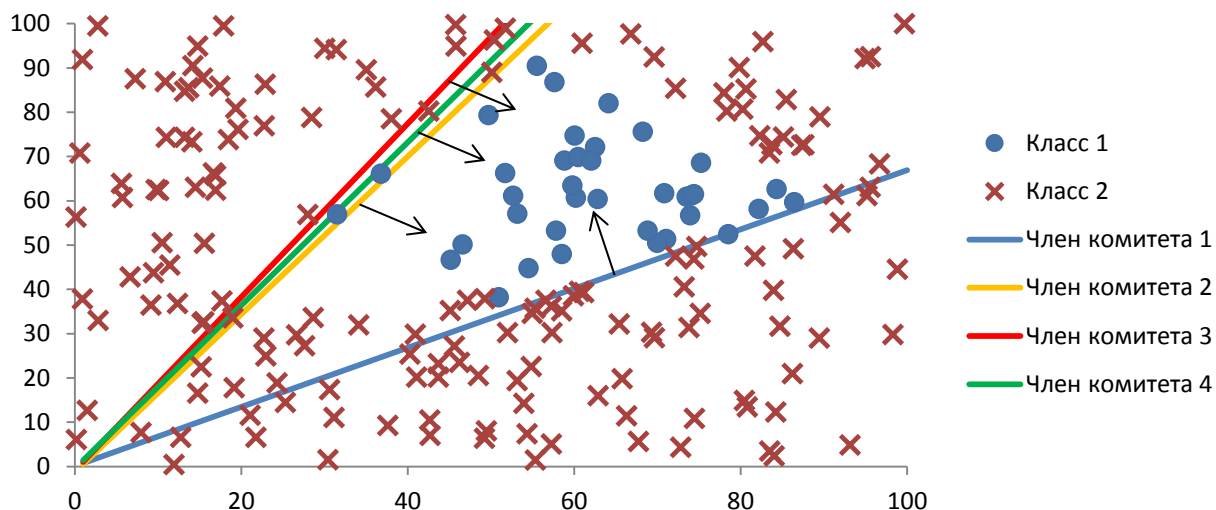


Рис. 26. Разделение множества комитетом из 4 членов, сгенерированным после 1000 операций мутаций и отбора.

Из рисунка 15 видно, что увеличение качества классификации после 1000 операций мутаций и отбора параметров комитета незначительно возросло по сравнению с 200 операциями мутаций и отбора. Так значение целевой функции составило 181 596, при дальнейших операциях мутаций и отбора не происходит увеличение качества классификации (уменьшение целевой функции). Как уже было отмечено выше, одной из основных проблем эволюционных алгоритмов является тенденция к сходимости алгоритма к локальному минимуму.

Заметим, что построение комитета через решение задачи частично-целочисленного программирования позволяет построить комитет единогласия из 4 членов однозначно классифицирующее множество, приведенное на рисунке 12. Параметры комитета, полученные за счет применения генетического алгоритма, можно использовать как начальное решение задачи частично-целочисленного программирования.

## Глава 3. Построение рейтинговой модели оценки заемщика на основе комитетных конструкций

### 3.1. Данные и оценка параметров заемщика

У нас имеются данные по 72 108 потребительским кредитам без обеспечения, выданным заемщикам мужчинам в период с 01.01.2011 по 01.08.2012, в различных территориях РФ, по которым было допущено 1 490 случаев дефолта (2,1%). Под дефолтом понимается возникновение просроченной задолженности по основному долгу или процентам свыше 90 дней. Заметим, что из генеральной совокупности исключены 138 заемщика со сроком просроченной задолженности свыше 330 дней, то есть заемщики, допустившие просроченную задолженность с первого платежа, поскольку, по мнению автора, дефолт данных заемщиков связан вероятнее всего с мошенническими операциями, проверка заемщиков на предмет возможного мошенничества должна осуществляться на этапе «фрод-мониторинга».

По заемщикам имеются следующие параметры:

1. Возраст заемщика;
2. Срок кредита;
3. Сумма кредита;
4. Доходы заемщика.

В таблицах 1 и 2 приведены средние значения параметров и стандартные отклонения в разрезе групп заемщиков.

Таблица 6. Средние значения параметров заемщиков.

Группы заемщиков	Возраст, лет	Срок кредита, лет	Сумма кредита, тыс. руб.	Доходы, тыс. руб.
Заемщики без дефолтов	38,20	3,98	235,10	41,04
Заемщики с дефолтом	30,90	3,81	201,30	34,07
<b>В среднем по выборке:</b>	<b>38,05</b>	<b>3,97</b>	<b>234,43</b>	<b>40,90</b>



Таблица 7. Стандартные отклонения параметров заемщиков.

Группы заемщиков	Возраст, лет	Срок кредита, лет	Сумма кредита, тыс. руб.	Доходы, тыс. руб.
Заемщики без дефолтов	11,16	1,28	186,61	16,11
Заемщики с дефолтом	10,25	1,56	208,16	16,39
<b>В среднем по выборке:</b>	11,19	1,28	187,13	16,15

Как видно из таблицы 3 в среднем заемщики со случаями дефолта моложе заемщиков без случаев дефолта. Заметим, что поскольку имеется логическая взаимосвязь (при этом данная зависимость является нелинейной - коэффициент корреляции составляет -1,6%) между доходами и возрастом заемщиков, то очевидно, что дефолтные заемщики имеют в среднем меньшие доходы. Из таблицы 4 следует, что у «дефолтных» заемщиков наблюдается больший разброс в сумме кредита и доходе.

Для подтверждения утверждения о зависимости возраста заемщика и доходов рассмотрим средние доходы заемщиков в разрезе возрастных групп.

Руб.

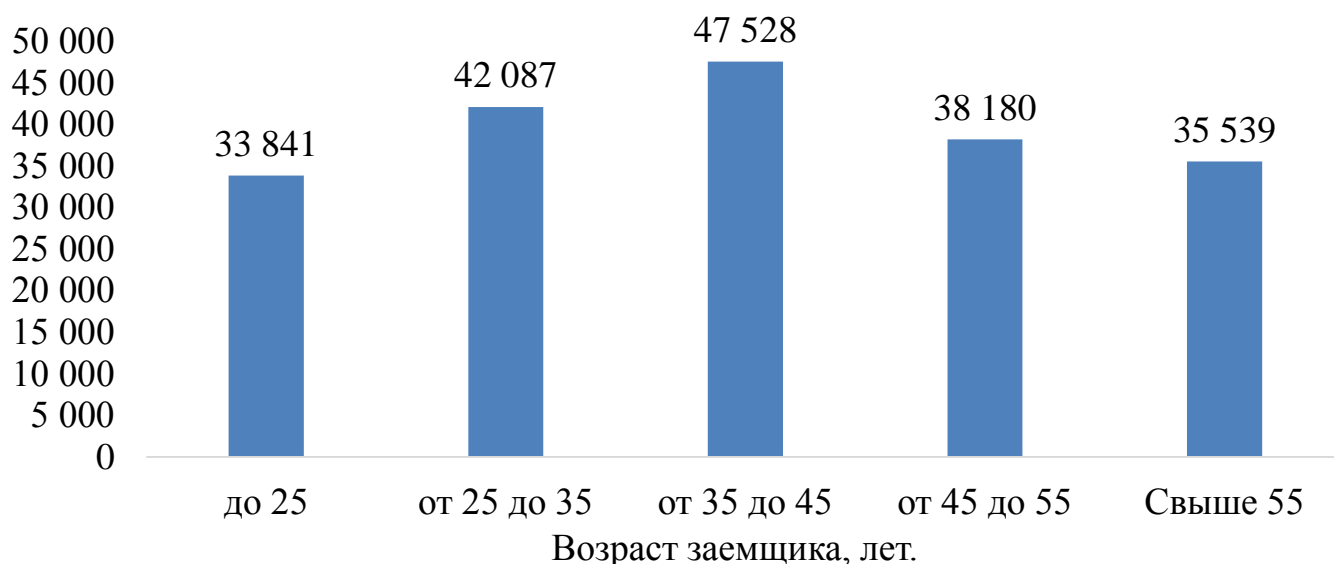


Рис. 24. Распределение доходов по возрастным группам.

Как видно из рисунка 24 наблюдается рост доходов до 45 лет, после 25 лет наблюдается снижение доходов заемщиков.

Далее рассмотрим линейную взаимозависимость параметров заемщиков. В таблице 3 приведены значения коэффициентов корреляции между параметрами.

Таблица 8. Линейные взаимозависимости параметров заемщиков

Параметр	Возраст	Срок кредита	Сумма кредита	Доходы
Возраст	100,0%	-16,2%	0,1%	-1,7%
Срок кредита	-16,2%	100,0%	41,3%	7,5%
Сумма кредита	0,1%	41,3%	100,0%	2,9%
Доходы	-1,7%	7,5%	2,9%	100,0%

В целом, исходя из коэффициентов корреляции, представленных в таблице 5, можно утверждать, что параметры заемщиков являются линейно независимыми. Наблюдается слабая линейная зависимость между сроком кредита и суммой кредита. Приведем гистограммы распределения параметров заемщиков.

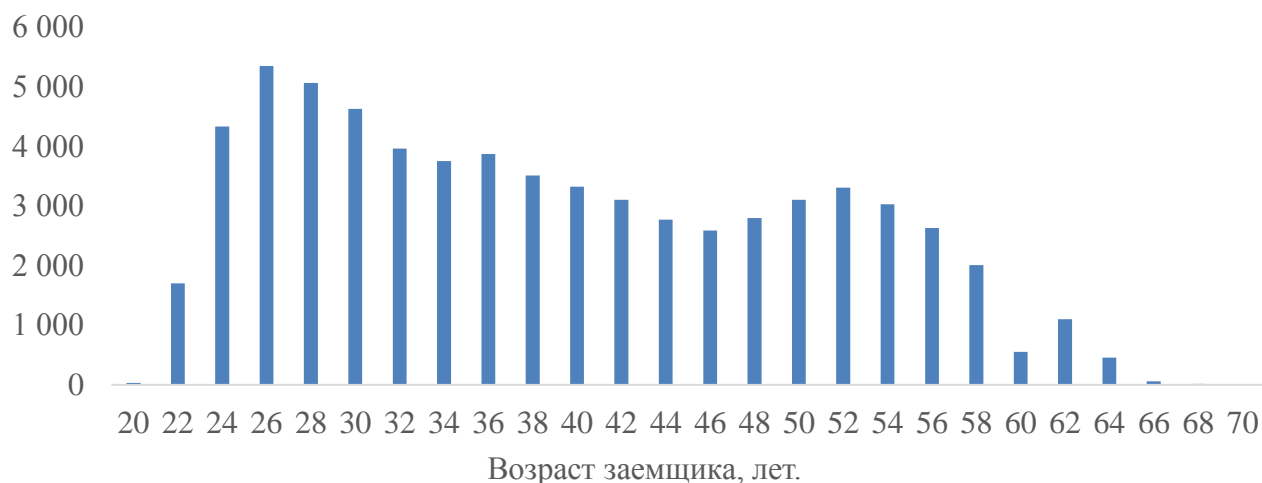


Рис. 25. Гистограмма распределения возрастов заемщиков.

Из рисунка 25 следует, что чаще всего заемщиками являются люди в возрасте от 24 до 36 лет, тогда как люди пенсионного и предпенсионного возраста менее склонны брать кредиты.

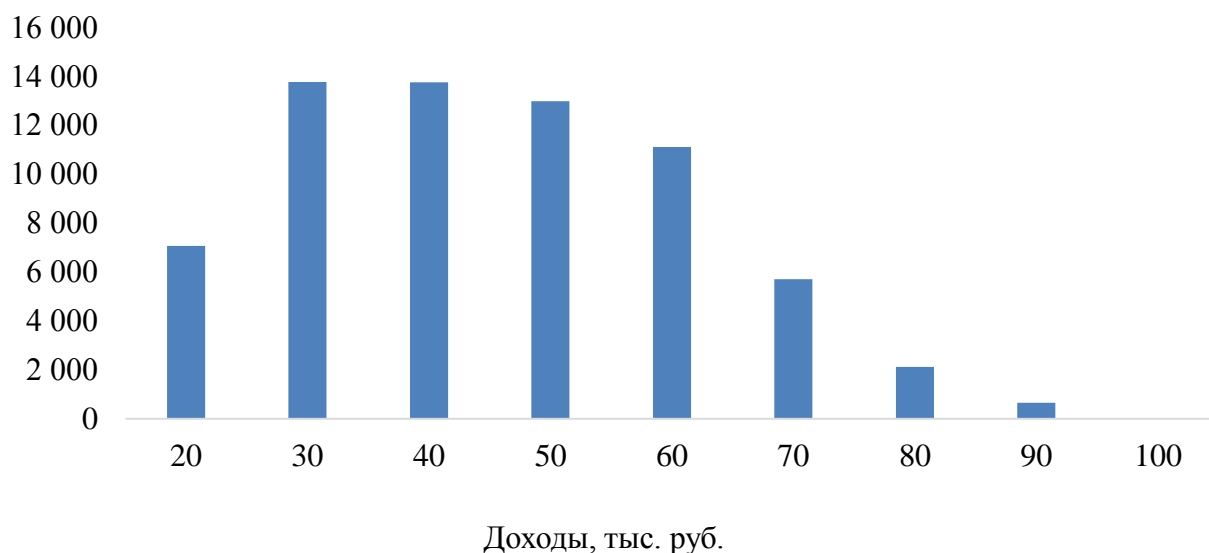


Рис. 26. Гистограмма распределения доходов заемщиков.

Как следует из рисунка 26, наибольшее число кредитов приходится на заемщиков с доходами от 20 до 50 тыс. руб. Доля заемщиков с доходами свыше 70 тыс. руб. незначительна.

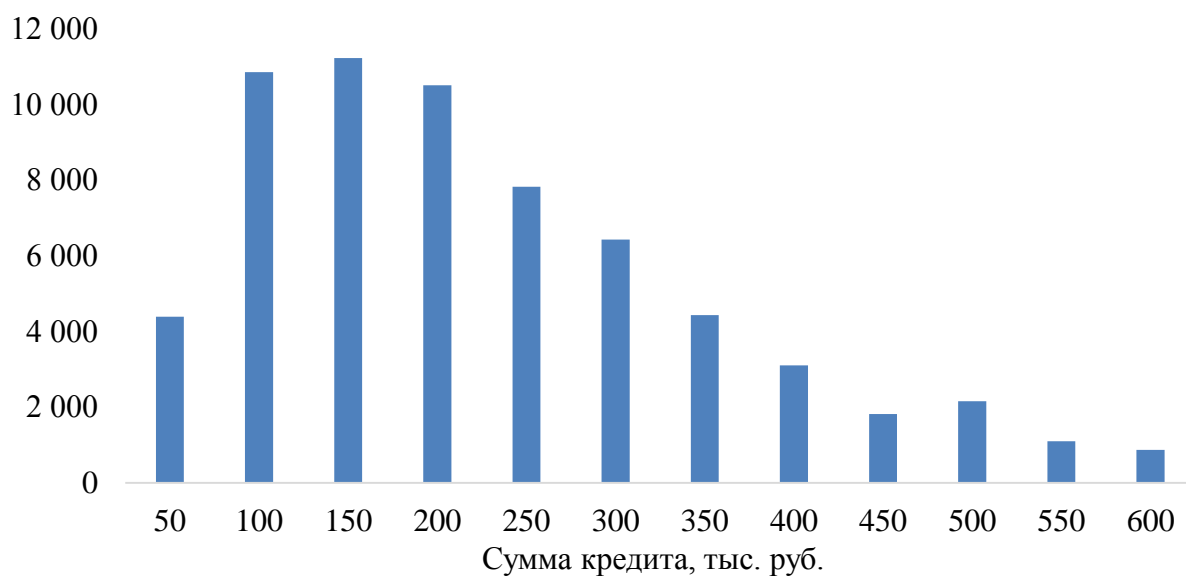


Рис. 27. Гистограмма распределения сумм кредитов.

Согласно рисунку 4, чаще всего заемщики берут кредиты от 50 до 200 тыс. руб. Доля кредитов с суммой свыше 400 тыс. руб. значительно меньше.

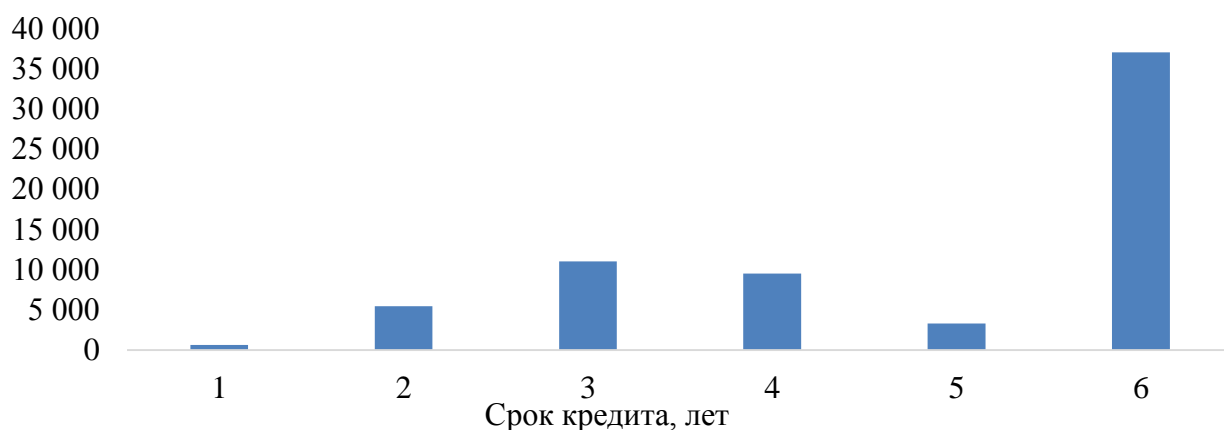


Рис. 28. Гистограмма распределения сумм кредитов.

Исходя из гистограммы, приведенной на рисунке 5 видно, что большая часть кредитов выдана на срок от 5 до 6 лет, следовательно, вероятнее всего параметр срок кредита будет иметь малое влияние на «дефолтность» заемщика.

Введем дополнительные параметры, исходя из экономической сути кредитного продукта, а именно, отношение суммы кредита к сроку кредита и отношение месячного платежа к доходам заемщика (месячный платеж рассчитывается как отношение суммы кредита к сроку кредита деленное на 12).

Рассмотрим средние значение и стандартное отклонение 2 новых введенных параметров.

Таблица 9. Средние значения параметров заемщиков.

Группы заемщиков	Среднее		Стандартное отклонение	
	Отношение месячного платежа к доходам, %	Отношение суммы кредита к сроку, тыс. руб./лет	Отношение месячного платежа к доходам, %	Отношение суммы кредита к сроку, тыс. руб./лет
Заемщики без дефолтов	14,36	6,68	13,84	5,51
Заемщики с дефолтом	17,40	6,91	24,83	6,63
<b>В среднем по выборке:</b>	<b>14,42</b>	<b>6,68</b>	<b>14,15</b>	<b>5,54</b>

Из таблицы 9 видно, что заемщики с дефолтами имеют значительно большее отношение месячного платежа к доходам, чем заемщики без дефолтов, следовательно, введение данного показателя позволит улучшить качество рейтинговой модели.

Поскольку нами введены дополнительные параметры имеет смысл провести проверку на мультиколлинеарность параметров.

Таблица 10. Линейные взаимозависимости параметров заемщиков

Параметр	Возраст	Срок кредита	Сумма кредита	Доходы	Отн. мес. платежа к доходам	Отн. суммы кредита к сроку
Возраст	100,0%	-16,2%	0,1%	-1,7%	10,3%	-31,9%
Срок кредита	-16,2%	100,0%	41,3%	7,5%	-2,6%	42,5%
Сумма кредита	0,1%	41,3%	100,0%	2,9%	62,6%	90,8%
Доходы	-1,7%	7,5%	2,9%	100,0%	-43,3%	1,7%
Отн. мес. платежа к доходам	10,3%	-2,6%	<b>62,6%</b>	<b>-43,3%</b>	100,0%	52,8%
Отн. суммы кредита к сроку	<b>-31,9%</b>	<b>42,5%</b>	<b>90,8%</b>	1,7%	<b>52,8%</b>	100,0%

Из таблицы 10 видно, что показатель отношение суммы кредита к сроку кредита имеет сильную корреляцию с большинством параметров заемщика. Параметр отношение месячного платежа к доходам имеет очевидную зависимость от суммы кредита и доходов заемщика. Поскольку выявлена сильная зависимость между параметрами имеет практический смысл произвести уменьшение параметров заемщика, для чего будет применен метод главных компонент. Уменьшение параметров заемщика позволит уменьшить время, требуемое для решения задачи.

### *3.1.1 Применение метода главных компонент для уменьшения размерности задачи*

Одним из способов уменьшить размерность изучаемой задачи является метод главных компонент. Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных. Иногда метод главных компонент называют преобразованием Кархунена — Лозва или преобразованием Хотеллинга (Hotelling transform). Компонента представляет собой линейную гиперплоскость, на которую осуществляется проецирование параметров. Уменьшение размерности задачи с минимальной потерей информации позволяет значительно

снизить вычислительную сложность задач, при этом избежав значительного снижения качества [98].

Покажем суть метода главных компонент на условном геометрическом примере. Пусть у нас имеется следующий набор наблюдений на плоскости.

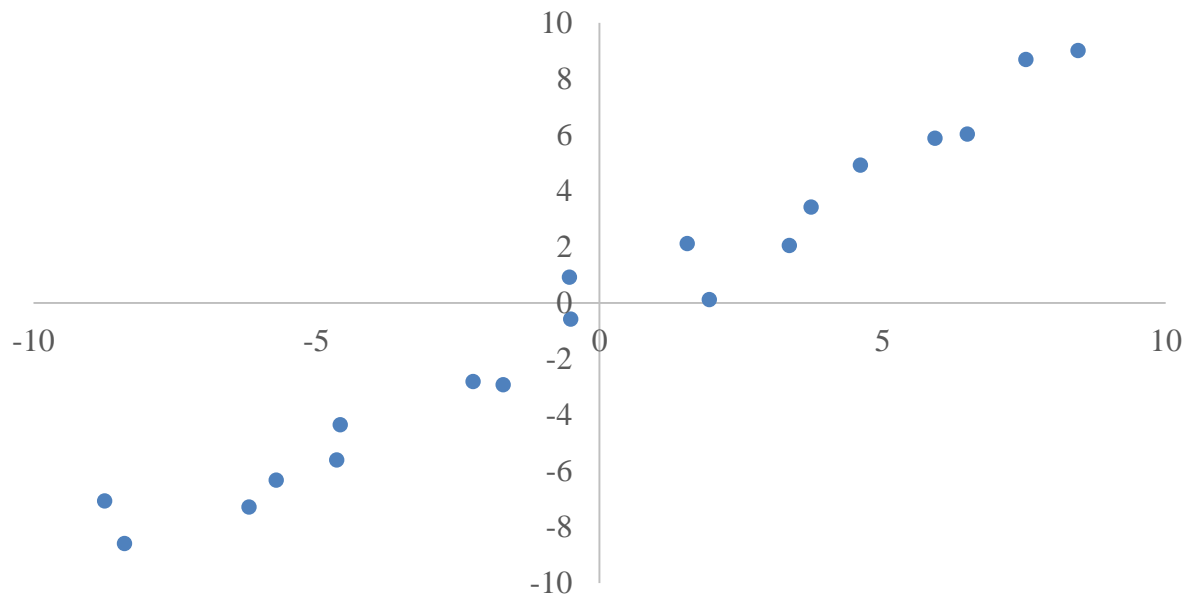


Рис. 29. Наблюдения в двухфакторном пространстве.

Первой компонентой (PC1) метода главных компонент является прямая, расположенная так чтобы вдоль нее происходило максимальное изменение данных. Проведем данную прямую.

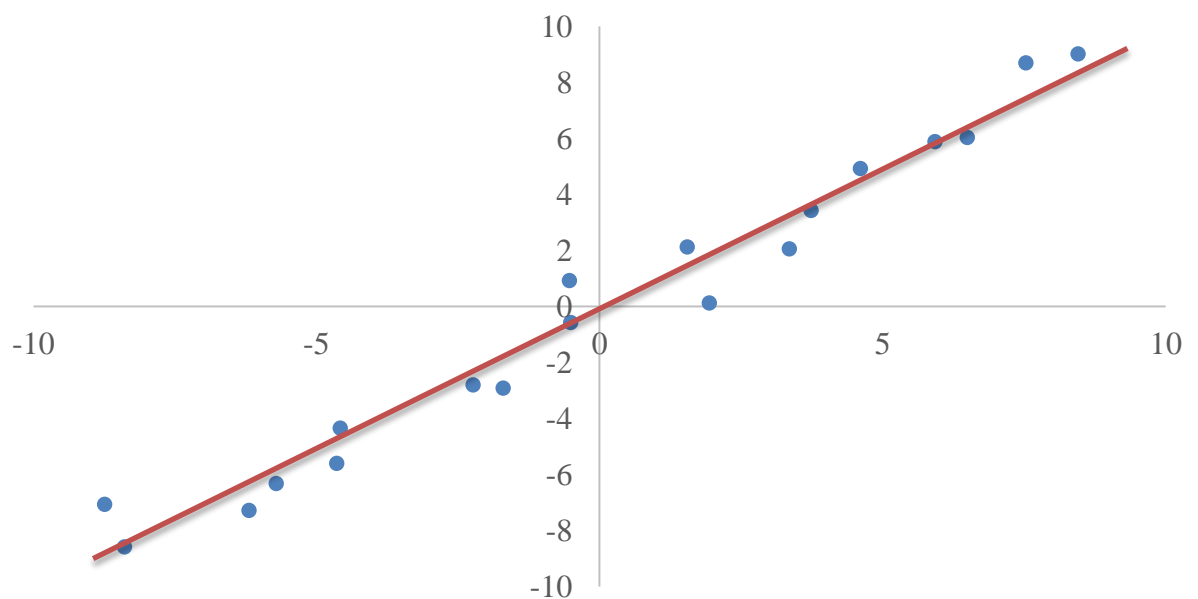


Рис. 30. Первая компонента PC1.

Предположением метода главных компонент является то, что отклонение данных от указанной прямой объясняется шумом, то есть ненужной информацией. Для проверки данного предположения осуществляется аналогичная операция с отклонениями от первой компоненты, то есть находится прямая, такая чтобы через нее происходило максимальное изменение отклонений от первой компоненты. Указанная операция осуществляется до того момента, пока шум уже не станет действительно шумом, т.е. случайным хаотическим набором величин [98].

В общем, многомерном случае, процесс выделения главных компонент происходит так:

1. Ищется центр облака данных, и туда переносится новое начало координат - это нулевая главная компонента (PC0)
2. Выбирается направление максимального изменения данных - это первая главная компонента (PC1)
3. Если данные описаны не полностью (шум велик), то выбирается еще одно направление (PC2) - перпендикулярное к первому, так чтобы описать оставшееся изменение в данных и т.д.

В результате осуществляется понижение размерности - переход от большого числа параметров к значительно меньшему, за счет чего снижается вычислительная сложность задач без потери качества решений [98].

Проведем анализ по методу главных компонент в программном пакете IBM SPSS Statistic 22. По итогам анализа выявлено 2 компоненты объясняющие параметры заемщиков.

Таблица 11. Компоненты объясняющие параметры заемщика

Параметры заемщика	Компонента	
	1	2
Возраст	-0,19	-0,49
Срок кредита	0,50	0,52
Сумма кредита	0,95	0,02
Доходы	-0,14	0,72
Отн. мес. платежа к доходам	0,70	-0,62

Отн. суммы кредита к сроку	0,95	0,18
-------------------------------	------	------

По итогам анализа по методу главных компонент были выявлены 2 компоненты, позволяющие с минимальной потерей качества снизить размерность решаемой задачи. То есть, был осуществлен переход от задачи с размерностью в 6 параметров к задаче с 2 параметрами заемщика.

Интересным будет рассмотреть, как распределены множества дефолтных и недефолтных заемщиков на плоскости.

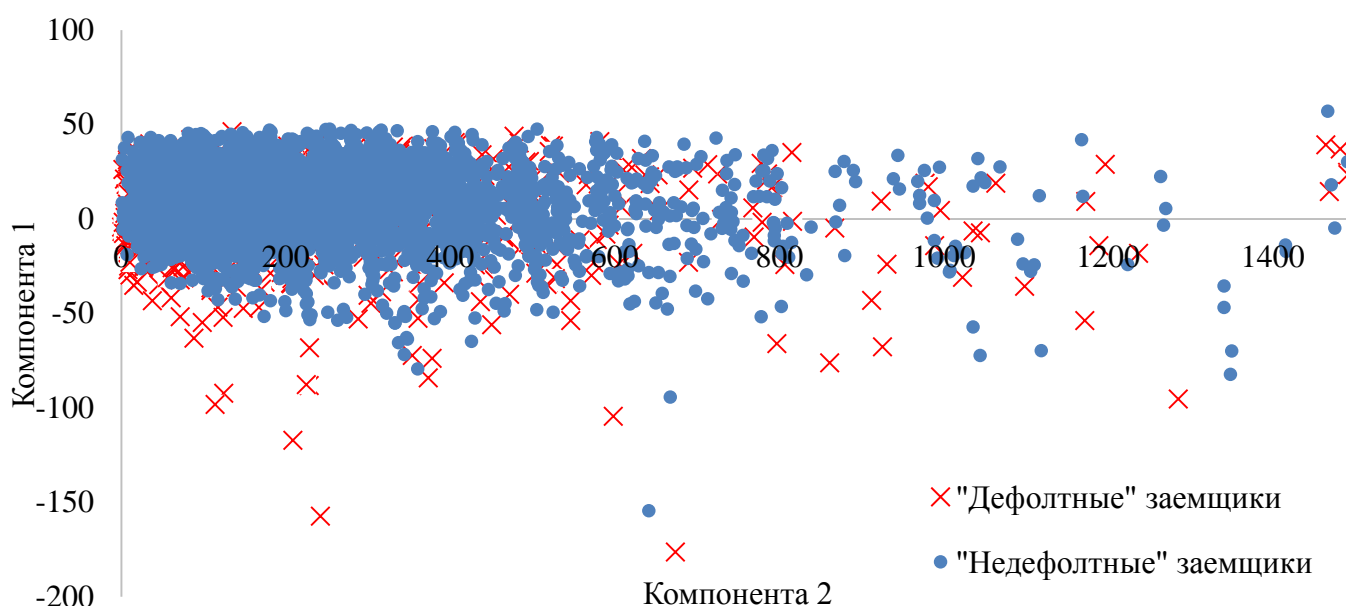


Рис. 31. Распределение заемщиков в пространстве 2 компонент.

Из рисунка 31 видно, что с удалением от начала координат наблюдается увеличение доли дефолтных заемщиков.

Заметим, что при анализе главных компонент нами использовалась выборка, на которой доля дефолтных заемщиков составляла 2%, возможно, что для более корректной оценки компонент необходимо увеличить долю дефолтных заемщиков до 50%, для этого увеличим выборку до 135 тысяч наблюдений путем копирования дефолтных заемщиков.

Увеличение доли дефолтных заемщиков привело изменению результатов метода главных компонент, что говорит о том, что корреляции в параметрах для дефолтных заемщиков отличаются от корреляций параметров недефолтных заемщиков. Были выделены 3 компоненты:



Таблица 12. Компоненты, объясняющие параметры заемщика после  
увеличения доли дефолтных заемщиков

Параметры заемщика	Компонента		
	1	2	3
Возраст	-0,13	-0,14	0,95
Срок кредита	0,57	0,46	-0,12
Сумма кредита	0,94	0,07	0,22
Доходы	-0,06	0,83	0,33
Отн. мес. платежа к доходам	0,61	-0,67	0,13
Отн. суммы кредита к сроку	0,95	0,12	-0,08

Попробуем рассмотреть группы заемщиков в пространстве 3 компонент, для этого воспользуемся пакетом Wolfram Mathematica.

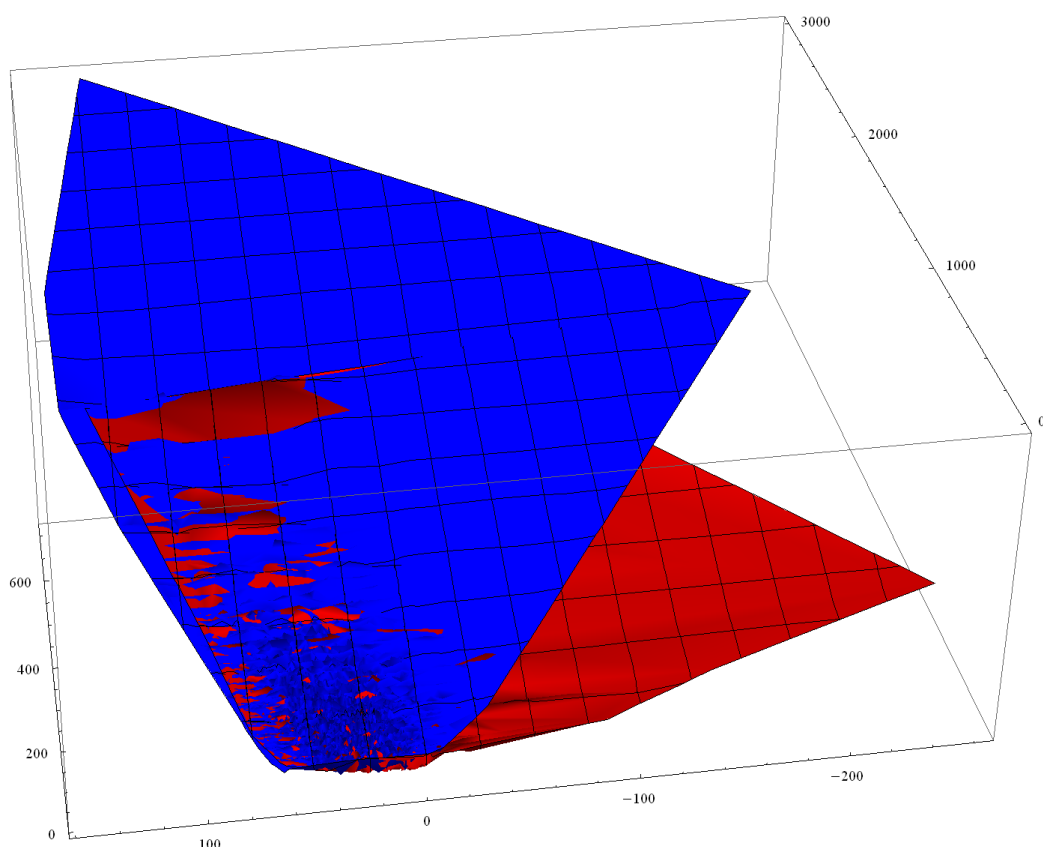


Рис. 32. Распределение заемщиков в пространстве 3 компонент.

На графике, приведенном на рисунке 32, по оси  $x$  (-200, 100) расположена 2 компонента, по оси  $y$  (0, 3000) расположена 1 компонента, по оси  $z$  (0, 600) – компонента 3. На графике множества «дефолтных» и «недефолтных» заемщиков

представлены как некоторые поверхности, где красная гиперповерхность – дефолтные заемщики, синяя - недефолтные. Как видно из графика увеличение числа компонентов привело к большей разделимости дефолтных и недефолтных заемщиков, следовательно увеличение доли дефолтных заемщиков на выборке было корректным.

В ходе дальнейшей работы будет произведено сравнение качества рейтинговых моделей, построенных с применением метода главных компонент и без применения метода главных компонент, что позволит оценить потери качества решения.

### **3.2. Построение рейтинговой модели оценки заемщика**

В главе 1 указано, что в рамках Международной конвергенции измерения капитала и стандартов капитала: новые подходы (Базель II) для оценки взвешенных по риску активов предусмотрены 2 подхода – стандартизированный и подход на основе внутренних рейтингов (Internal Ratings-Based Approach (IRB Approach)).

При применении подхода на основе внутренних рейтингов кредитной организацией каждому заемщику присваивается соответствующий рейтинг в зависимости от вероятности дефолта по кредиту. Чаще всего для построения рейтинговой модели используется Logit метод, позволяющий на основании параметров заемщика определить вероятность возникновения дефолта по кредиту. Далее, значения Logit модели группируются в зависимости от уровня кредитного риска, так например, если Logit модель принимает значения от 0% до 2%, то заемщику присваивается 1 рейтинг. В дальнейшем по каждой рейтинговой группе (рейтингу) определяется средний уровень вероятности дефолта и средний уровень потерь в случае дефолта, заметим, что для определения вероятности дефолта и уровня потерь в случае дефолта могут применяться различные рейтинговые модели. Группировка заемщиков по рейтингам во многом связана с дальнейшим расчетом взвешенных по риску активов и методами валидации рейтинговых моделей.

В данной части работы будет рассмотрено применение метода комитетов к построению рейтинговой модели.

### *3.2.1. Построение рейтинговой модели с применением метода комитетов*

Как уже было отмечено ранее, метод комитетов выдает бинарное решение (1 – дефолт, 0 – отсутствие дефолта). Бинарный ответ комитета не подходит для построения рейтинговой модели, поскольку в данном случае к нас имелось бы только 2 рейтинга заемщиков, что недостаточно для применения IRB-подхода.

Заметим, что в ходе построения комитета происходит построения членов комитета, решения которых в дальнейшем преобразуются в зависимости от комитетного правила, так например в комитете большинства, в случае, если более половины членов комитета принимают решение, что по кредиту возникнет дефолт, комитет выдает решение о возникновении дефолта. В рамках данной работе нами применяется  $p$ -комитет с произвольным правилом комитетного решения, которое определяется исходя из значений переменной  $p$  (см. формулы ранее).

Поскольку при решении задачи целочисленного линейного программирования для построении комитета используется целевая функция, минимизирующая одновременно и ошибки классификации комитета и ошибки классификации каждого из членов комитета, то можно выдвинуть предположение, что с увеличением числа членов, «голосующих» за возникновение дефолта по кредиту, вероятность дефолта будет возрастать.

Проверим данное предположение на фактических данных, полученных от кредитной организации.

Для построения комитета сформируем обучающую выборку составляющую около 10% от генеральной совокупности (7 000 заемщиков без случаев дефолта и 140 заемщиков со случаями дефолта). Для классификации заемщиков построим комитет, состоящий из 7 членов. Время, затраченное на построение комитета, составило около 16 часов. Протестируем результаты комитетного решения на

выборке, состоящей из 67 458 заемщиков без случаев дефолта и 1 340 заемщиков со случаями дефолта.

Результаты комитетного решения приведены в таблице 4.

Таблица 13. Результаты комитетного решения

Голосование комитета	Наблюдения со случаями дефолта	Наблюдения без случаев дефолта	Доля заемщиков в дефолте
«За» дефолт	1 930	47 842	4,0%
«Против» дефолта	375	43 537	0,9%
Итого	1 340	67 458	2,0%

Комитетное правило определенное, при построении комитета можно сформулировать следующим образом - в случае, если более 1 членов комитета, голосуют за дефолт, то комитет принимает решение о возникновении дефолта.

Рассмотрим результаты комитетного решения в зависимости от числа членов проголосовавших за дефолт.

Таблица 14. Вероятность дефолта в зависимости от числа проголосовавших членов

Число членов комитета «проголосовавших» за дефолт	Число договоров	Число договоров в дефолте	PD группы	Доля группы в тестовой выборке
0	43509	391	0,9%	64,5%
1	2901	50	1,7%	4,3%
2	1214	30	2,5%	1,8%
3	3171	76	2,4%	4,7%
4	13087	549	4,2%	15,4%
5	3036	218	7,2%	4,5%
6	540	26	4,9%	4,8%
Итого:	67458	1340	2,0%	100%

Результаты, приведенные в таблице 14, подтверждают предположение, о возрастании вероятности дефолта с увеличением числа членов комитета, проголосовавших за дефолт. Заметим, что вероятность дефолта в случае, если за дефолт проголосовало 6 члена, меньше, чем если за дефолт проголосовало 6

члена, что возможно связано, с малым числом наблюдений в тестовой выборке за которые голосуют ровно 6 членов. Заемщики, за дефолт которых «проголосовало» бы свыше 6 членов комитета, в тестовой выборке отсутствуют. Поскольку с возрастанием числа членов комитета возрастает вероятность дефолта по кредиту, можно выделить рейтинговые группы.

Таблица 15. Рейтинговые группы заёмщиков

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	PD группы	Доля группы в тестовой выборке
1	0	0,9%	64,5%
2	1	1,7%	4,3%
3	2-3	2,4%	6,5%
4	4	4,2%	15,4%
5	5 и более	6,0%	9,3%
Итого:		2,0%	100%

Как видно из таблицы 15 комитетом из 7 членов было выделено 5 групп заемщиков, с возрастанием PD в зависимости от числа «проголосовавших» за дефолт членов комитета. Заметим, что в рейтинговую группу 3 входят заемщики, за которых проголосовали 2 или 3 члена комитета, что связано с тем, что доли заемщиков в дефолте у данных групп близки, так же заемщики, за которых проголосовали 5 или 6 членов комитета объединены в одну группу, поскольку доля дефолтов у заемщиков за которых проголосовало 6 членов комитета меньше, чем в группе, за которую проголосовало 5 членов.

Рассмотрим возможность увеличения качества рейтинговой модели за счет использования матриц миграций.

### 3.2.2. Последовательная оптимизация комитетного решения

Как было отмечено выше построение комитета осуществляется на обучающей выборке, содержащей 10% наблюдений генеральной совокупности. Построение комитета на выборке, содержащей только 10%, связано с тем, что построение комитета на большем объеме обучающей выборке затруднительно в связи с возрастанием числа целочисленных переменных. Для увеличения

качества комитетного решения и, соответственно, качества рейтинговой модели можно применить метод последовательной оптимизации для построения комитета на обучающей выборке, содержащей большее число наблюдений.

Последовательная оптимизация комитетного решения заключается в фиксации параметров части членов комитета и решения задачи частично-целочисленного программирования на выборке содержащей большее число наблюдений.

Так, например, для построения комитета из 7 членов на выборке из 7 140 наблюдений (7 000 наблюдений без случаев дефолта, 140 наблюдений со случаями дефолта) необходимо решить задачу частично-целочисленного программирования из 57 121 целочисленных переменных. После построения комитета фиксируются параметры 6 членов из 7, а обучающая выборка увеличивается до 14 280 (14 000 наблюдений без случаев дефолта, 280 наблюдений со случаями дефолта), и решается задача частично целочисленного программирования, содержащая 27 560 целочисленных переменных. Указанная операция последовательно осуществляется для каждого из членов комитета. Также возможна фиксация не членов комитета, а параметров членов комитета, так, например, можно зафиксировать 5 из 6 параметров, решение данной задачи, несмотря на значительное число целочисленных переменных, занимает меньшее время поскольку уменьшается область допустимых значений.

Рассмотрим изменение качества рейтинговой модели за счет применения последовательной оптимизации. Приведем параметры рейтинговой модели исходя из комитетного решения из 7 членов, полученного на обучающей выборке, состоящей из 7000 наблюдений без случаев дефолта и 130 наблюдений со случаями дефолта (10% от генеральной совокупности).

Таблица 16. Параметры рейтинговой модели.

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	PD группы	Доля группы в тестовой выборке
1	0	0,8%	64,5%

2	1	1,5%	4,3%
3	2-3	2,1%	6,5%
4	4	3,8%	15,4%
5	5 и более	5,4%	9,3%
Итого:		1,8%	100,0%

Далее зафиксируем параметры 6 членов и увеличим обучающую выборку до 14 000 наблюдений без случаев дефолта и 260 наблюдений со случаями дефолта (20% от генеральной совокупности).

Важным при последовательно оптимизации комитетного решения является отбор членов комитета, которые будут зафиксированы, и которые будут изменяться. В данной работе последовательно оптимизировались члены комитета имеющие минимальную точность на обучающей выборке. Так же при выборе членов комитета необходимо учитывать совпадение решений с другими членами, так если комитет член комитета принимает решения во многом схожие с решениями другого члена, то целесообразно провести последовательную оптимизацию обоих членов, предварительно зафиксировав решения других членов.

Таблица 17. Параметры рейтинговой модели, после оптимизации одного  
из членов.

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	PD группы	Доля группы в тестовой выборке
1	0	0,7%	55,6%
2	1	1,5%	10,6%
3	2	2,2%	11,4%
4	3	3,2%	13,0%
5	4	5,6%	8,6%
6	5 и более	10,3%	0,8%
Итого:		1,8%	100%

Для сравнения качества рейтинговых моделей нами используется мера AR (Accuracy Ratio), другое ее название – индекс Джини (GINI index, GINI). AR

вычисляется как отношение площадей полученной модели к «идеальной модели», определяемых с помощью кривой CAP (cumulative accuracy profile). Пусть все наблюдения отсортированы по присвоенному скоринговому баллу от наихудшего к наилучшему. Тогда точки CAP кривой определяются соотношением доли наблюдений к такой доле «плохих» наблюдений, где скоринговый балл хуже, чем лучший балл любого наблюдения, включенного в последнюю группу наблюдений. Применение меры AR производится на этапе валидации рейтинговой модели, подробнее методы валидации рейтинговых моделей рассмотрены в Главе 3.

Так значение AR для рейтинговой модели, приведенной в таблице 16, составило 42,4%, для рейтинговой модели, приведенной в таблице 17 – 43,0%, следовательно, применение последовательной оптимизации позволило увеличить качество рейтинговой модели.

Далее проведем последовательную оптимизацию параметров членов комитета и членов комитета, последовательная оптимизация осуществляется до того момента, когда точность комитетного решения перестает улучшаться. Полученная рейтинговая модель приведена в таблице 18.

Таблица 18. Параметры рейтинговой модели, после последовательной оптимизации всех членов комитета.

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	PD группы	Доля группы в тестовой выборке
1	0	0,7%	54,3%
2	1	1,5%	10,1%
3	2	2,1%	9,6%
4	3	3,0%	15,8%
5	4	5,2%	8,9%
6	5 и более	10,3%	1,3%
Итого:		1,8%	100%

Значение меры AR для полученной рейтинговой модели составило 43,4%.

Заметим, что, несмотря на то, что метод последовательной оптимизации является эвристическим, применение указанного метода позволяет увеличить



качество комитетного решения и, соответственно, рейтинговой модели, в связи с чем, по нашему мнению, при решении практических задач классификации заемщиков, применение последовательной оптимизации комитетного решения целесообразно.

### *3.2.3. Динамика вероятности дефолта кредитов физических лиц в период со 2 квартала 2011 по 2 квартал 2014*

В главе 1 как один из способов расчета кредитного риска были приведены матрицы миграций. Рассмотрим динамику вероятности дефолта кредитов физических лиц в период со 2 квартала 2011 по 2 квартал 2014 годов, рассчитанную на основании матриц перехода / миграций. Матрица переходов / миграций позволяет оценить вероятность дефолта заемщика в течение 12 месяцев исходя из вероятностей перехода заемщиков по группам просрочки (без просроченных платежей, с просроченными платежами от 1 до 30 дней, от 31 до 60 дней, от 61 до 90 дней) в течение месяца.

Расчет матриц миграций производился по данным раздела 3 «Информация по сгруппированным в портфели однородным требованиям и ссудам, предоставленным физическим лицам» формы 0409115 «Информация о качестве активов кредитной организации (банковской группы)» кредитных организаций, зарегистрированных на территории г. Москвы. Форма 0409115 ««Информация о качестве активов кредитной организации (банковской группы)» утверждена Указанием Банка России от 12.11.2009 N 2332-У «О перечне, формах и порядке составления и представления форм отчетности кредитных организаций в Центральный банк Российской Федерации».

По данным форм 0409115 были построены матрицы миграций и получены вероятности дефолта ссуд без просроченных платежей по ипотечным кредитам, по автокредитам и по иным потребительским кредитам за период со 2 квартала 2011 по 2 квартал 2014 года. На рисунке 33 приведены средние значения вероятностей дефолта за квартал.

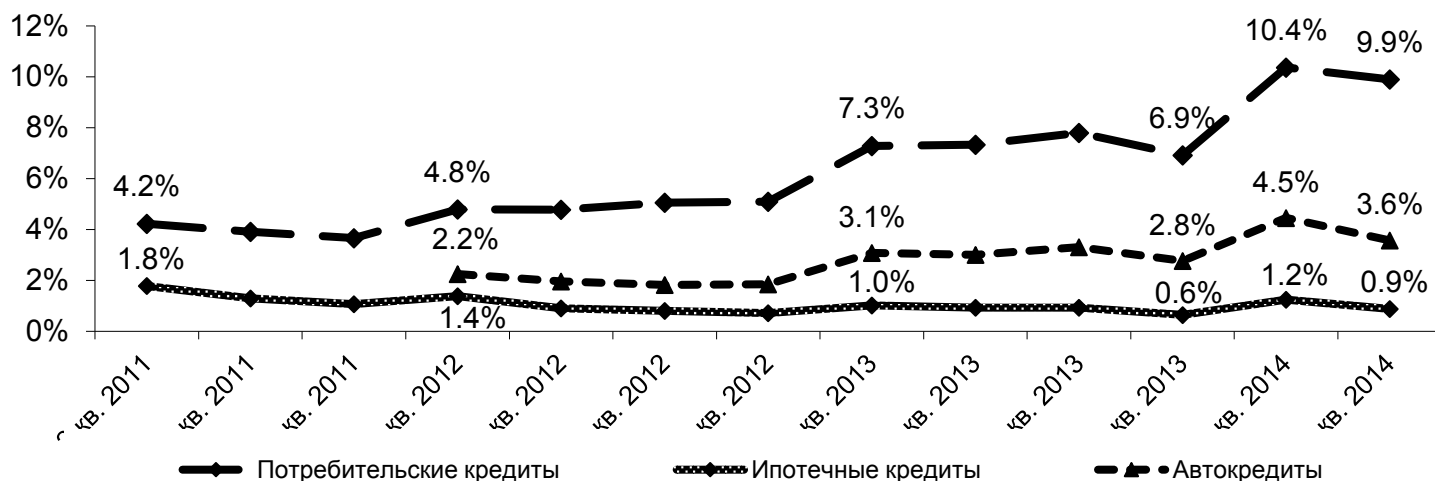


Рис. 33. Динамика вероятности дефолта по кредитам физических лиц

Как видно из рис. 33, в 1 квартале 2014 года произошло значительное увеличение вероятности дефолта по потребительским кредитам по сравнению с 4 кварталом 2013 года (рост на 3,5 п.п.), в целом же за период со 2 квартала 2011 по 2 квартал 2014 вероятность дефолта увеличилась более чем в 2 раза (с 4,2% до 9,9%). При этом по ипотечным кредитам наблюдается противоположная динамика – за аналогичный период вероятность дефолта снизилась с 1,8% до 0,9%. Рост вероятности дефолта в 1 полугодии 2014 года по потребительским кредитам привел к тому, что кредитные организации с низким уровнем диверсификации ссудного портфеля (существенной долей вложений в потребительские кредиты) вынуждены были сформировать дополнительные резервы на возможные потери, что привело к снижению прибыли.

Проведем сравнение совокупных резервов, сформированных в соответствии с Положением Банка России № 254-П от 26.03.2004 «О порядке формирования кредитными организациями резервов на возможные потери по ссудам, по ссудной и приравненной к ней задолженности» (далее – Положение №254-П) [3], и резервов, рассчитанных по IRB-подходу в размере ожидаемых потерь по портфелю. Для расчета ожидаемых потерь по портфелю необходимо дополнительно вычислить вероятность дефолта по кредитам с просрочкой от 1 до 30 дней, от 31 до 60 дней, от 61 до 90 дней, по кредитам с просрочкой свыше 90 дней вероятность дефолта приравнивается к 100%.

В рамках IRB подхода Базеля II предполагается, что резервы, формируемые кредитной организацией, не должны быть меньше ожидаемых потерь по портфелю, иначе разница между сформированными резервами и ожидаемыми потерями должна вычитаться из капитала банка (§43, Базель II). Уровень ожидаемых потерь (EL) по розничным кредитам рассчитывается по следующей формуле:  $EL = PD * LGD$ , где PD – вероятность дефолта по ссуде, LGD – уровень потерь банка в случае дефолта, ожидаемые потери по портфелю соответственно равняются произведению EL на EAD, где EAD – величина кредитного требования, подверженная риску дефолта (средства, предоставленные заемщику) (§375-376, Базель II [2]). Заметим, что по действующим формам отчетности кредитных организаций не представляется возможным рассчитать потери в случае дефолта, в связи с чем нами предполагается, что потери по иным потребительским кредитам в случае дефолта в среднем составляют 45%, по ипотечным кредитам - 25% (значения потерь в случае дефолта по ипотечным кредитам и потребительским кредитам, приведены в приложении 3 к Базелю II).

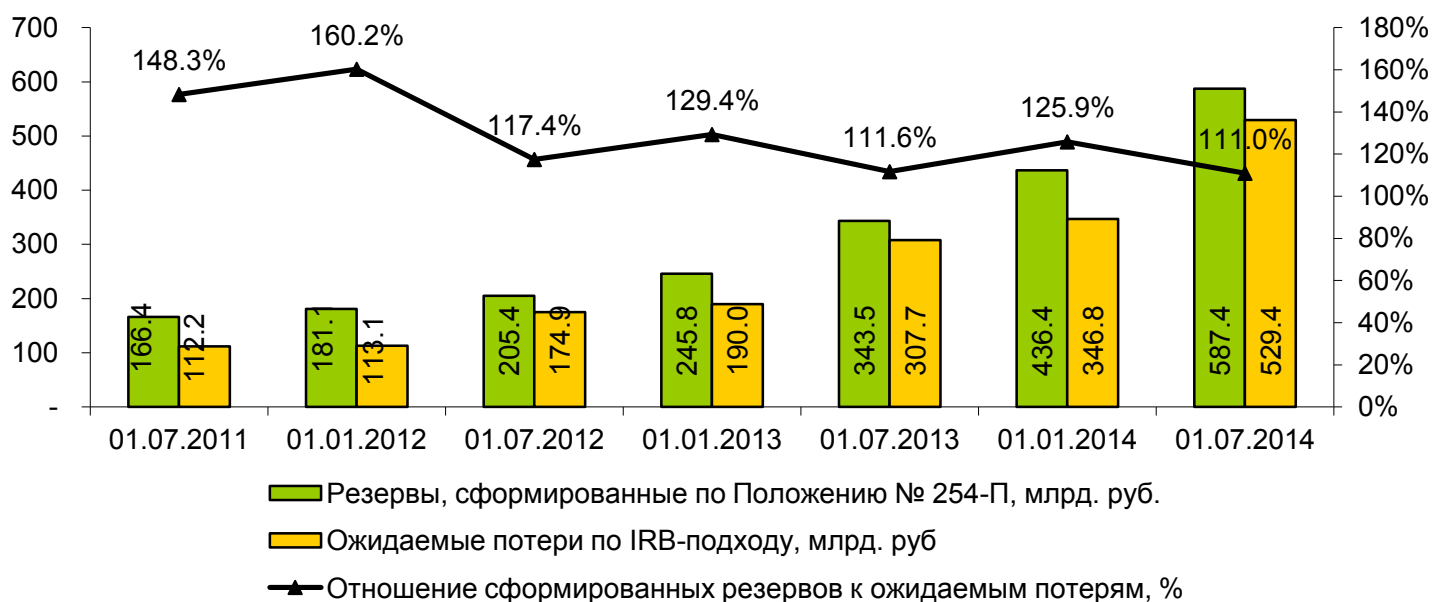


Рис. 34. Отношение резервов, сформированных по Положению № 254-П, к ожидаемым потерям по портфелю потребительских кредитов.

Как видно из рисунка 34, в периоде со 2 квартала 2011 по 2 квартал 2014 отмечается тенденция к снижению отношения сформированных резервов и

ожидаемых потерь по потребительским кредитам, рассчитанных в рамках IRB-подхода, с 148,2% до 111,0%.

Отметим, что в расчетах предполагается, что по потребительским кредитам потери в случае дефолта (LGD) составят 45%, что может не соответствовать реальным потерям в случае дефолта [92]. По нашему мнению средний уровень потерь по потребительским кредитам, в связи с высокой долей необеспеченных кредитов, составляет от 70-80%. Увеличение потерь в случае дефолта на 5 п.п. приводит к увеличению ожидаемых потерь по портфелю потребительских кредитов на 58,8 млрд. руб. При ожидаемых потерях в случае дефолта в 50% сформированные по Положению №254-П резервы будут близки к ожидаемым потерям по портфелю.

На рисунке 18 приведены результаты сравнения резервов, сформированных по Положению № 254-П, и ожидаемых потерь (расчетных резервов) по IRB-подходу по ипотечным кредитам.

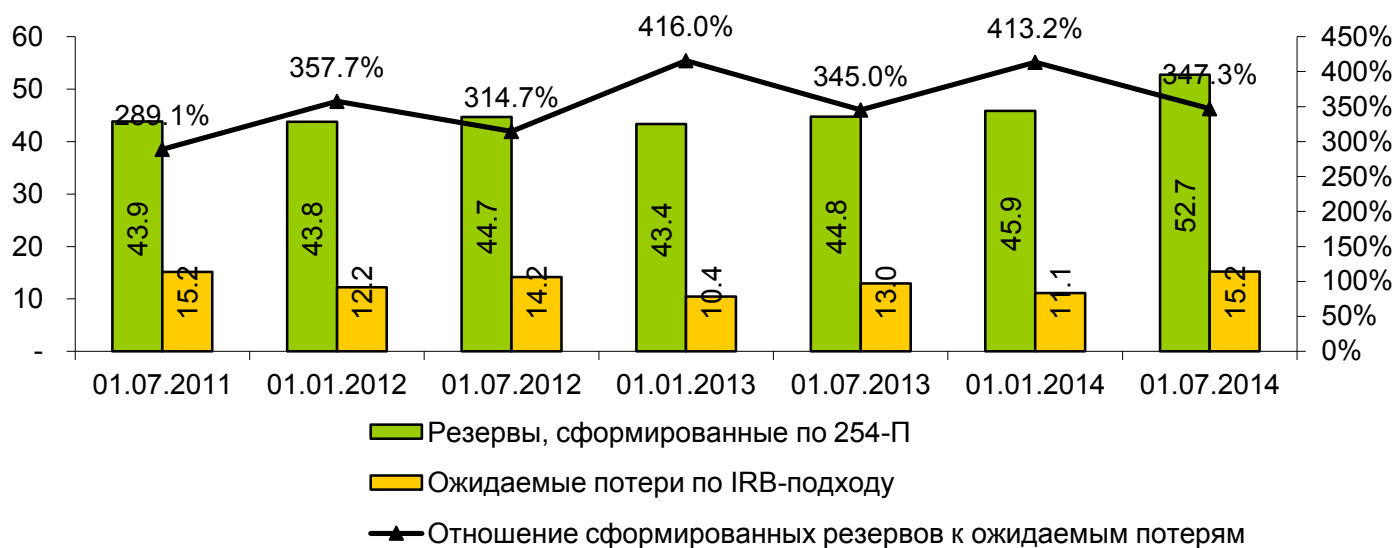


Рис. 35. Отношение резервов, сформированных по Положению № 254-П, к ожидаемым потерям по портфелю ипотечных кредитов.

Как видно из рисунка 35, в периоде со 2 квартала 2011 по 2 квартал 2014 отмечается тенденция к росту отношения сформированных резервов и ожидаемых потерь, рассчитанных в рамках IRB-подхода, с 289,1% до 347,3%.

Отметим, что в расчетах предполагается, что потери по ипотечным кредитам в случае дефолта составят 25%, увеличение потерь в случае дефолта на

5 п.п. приводит к увеличению ожидаемых потерь по портфелю ипотечных кредитов на 3,0 млрд. руб. При ожидаемых потерях в случае дефолта в 87% сформированные резервы будут соответствовать ожидаемым потерям по портфелю.

За период с 4 квартала 2013 по 2 квартал 2014 значительно увеличилась вероятность дефолта по потребительским кредитам (с 6,4% до 9,9%), что привело к формированию дополнительных резервов на возможные потери, при этом вероятность дефолта по ипотечным кредитам увеличилась незначительно (с 0,6% до 0,9%). Формирование дополнительных резервов на возможные потери в связи с ухудшением качества кредитного портфеля привело к значительному снижению прибыли за 1 полугодие 2014 кредитных организаций, ориентированных на потребительское кредитование. За период со 2 квартала 2011 по 2 квартал 2014 вероятность потерь по потребительским кредитам увеличилась с 4,2% до 9,9%, в случае сохранения тенденции к увеличению вероятности дефолта можно ожидать дальнейшего снижения прибыли кредитных организаций, ориентированных на потребительское кредитование. В целом, если сравнивать объемы сформированных резервов по Положению № 254-П с ожидаемыми потерями по IRB подходу Базеля II, то по потребительским кредитам отмечается тенденция к снижению отношения резервов по Положению № 254-П к ожидаемым потерям по IRB подходу Базеля II, тогда как по ипотечным кредитам отмечается противоположная тенденция. Заметим, что без точного расчета потерь в случае дефолта (LGD) не представляется возможным утверждать, достаточность резервов, сформированных по Положению №254-П, для покрытия ожидаемых потерь, рассчитанных по IRB подходу Базеля II. По нашему мнению, для поддержания покрытия резервами на уровне, превышающем ожидаемые потери, целесообразно увеличение нормативов резервирования по Положению №254-П по потребительским кредитам, при этом по кредитам обеспеченным жильем, является допустимым снижение нормативов резервирования, в связи со значительным превышением резервов над ожидаемыми потерями.

### 3.2.4. Применение матриц миграций для увеличения качества рейтинговой модели.

Ранее заемщики были разделены нами 2 класса – заемщики со случаями дефолта и заемщики без случаев дефолта, и была построена рейтинговая модель определяющая вероятность дефолта, исходя из числа членов, которые проголосовали за дефолт. Рассмотрим возможность построения рейтинговых моделей, которые будут определять вероятности перехода кредитного договора из одной группы просрочки в другую в течение одного месяца.

Нами были проанализированы переходы кредитов, выданных с 01.02.2011 по 01.08.2012, из одной группы просрочки в другую в течение месяца за период с 01.02.2012 по 01.08.2013 (общее число кредитов – 72 052). В таблицах 19 и 20 приведены данные о состоянии кредитов на начало периода и числе переходов из одной группы просрочки другую за рассматриваемый период.

Таблица 19. Данные о числе кредитов в группах на начало периода

-	Без просроченной задолженности на начало месяца	С просроченной задолженностью от 1 до 30 дней на начало месяца	С просроченной задолженностью от 31 до 60 дней на начало месяца	С просроченной задолженностью от 61 до 90 дней на начало месяца
Всего за анализируе мый период	1 073 961	21 913	5 867	3 850

Таблица 20. Данные о числе переходов из группы в группу в течение периода

-	Выход на просрочку в течение месяца	Переход на срок от 31 до 60 дней в течение месяца	Переход на срок от 61 до 90 дней в течение месяца	Переход на срок свыше 90 дней в течение месяца
Всего за анализируе мый период	17 255	6 096	4 080	3 056

Из таблиц 19 и 20 видно, что в течение анализируемого периода общее число кредитов, находящихся в группе без просроченных платежей составило 1 073 961 единиц, при этом число выходов на просрочку составило 17 255, средняя частота выхода на просрочку – 1,6%. Заметим, что ранее генеральная совокупность состояла из 74 458 кредитов (из них 1 480 в дефолте). По нашему мнению, построение рейтинговой модели на большем числе наблюдений позволит получить более статистически точные результаты.

Исходя из данных таблиц 19 и 20 построим матрицу миграций. Матрица переходов/миграций представляет собой матрицу частот (вероятностей) перехода из одного состояния в другое (размерность матрицы составляет  $n * p$ , где  $n$  – число групп,  $p$  – число возможных состояний). Как уже было отмечено ранее в главе 1 для каждого кредитного договора без просроченной задолженности на начало месяца существует только 2 возможных состояния на конец месяца – отсутствие просроченной задолженности по договору или наличие просрочки по договору. То есть, если 5,0% кредитных договоров вышли на просрочку, то у 95,0% не возникло просрочки. Соответственно для кредитных договоров с просроченной задолженностью от 1 до 30 дней на начало месяца существует 2 возможных состояния на конец месяца – отсутствие просроченной задолженности или наличие просроченной задолженности от 31 до 60 дней.

В таблице 21 приведена матрица переходов/миграций кредитных договоров за месяц, построенная по данным за период с 01.08.2012 по 01.08.2013.

Таблица 21. Матрица миграций кредитных договоров.

		Группа на конец месяца				
		Без просрочки	С просрочкой от 1 до 30 дней	С просрочкой от 31 до 60 дней	С просрочкой от 61 до 90 дней	С просрочкой свыше 90 дней
Группа на начало месяца	Без просрочки	98,4%	1,6%	0,0%	0,0%	0,0%
	С просрочкой от 1 до 30 дней	72,2%	0,0%	27,8%	0,0%	0,0%
	С просрочкой от 31 до 60 дней	30,5%	0,0%	0,0%	69,5%	0,0%
	С просрочкой от 61 до 90 дней	20,6%	0,0%	0,0%	0,0%	79,4%

В приложении приведены подробные данные о частоте переходов из группы в группу в течение месяца.

Вероятность дефолта согласно матрице миграций, приведенной в таблице 21 составила 1,9%, заметим, что средняя вероятность дефолта среди анализируемых кредитных договоров составила 2,0%. Расчет вероятности дефолта проводился в соответствии с динамической моделью, предложенной в главе 1.

Интересным будет рассчитать матрицы миграций и вероятности дефолта для каждой из рейтинговых групп, полученных на предыдущем этапе. В таблице 22 приведены данные о частоте переходов из групп с более низким сроком просроченной задолженности в группу с более высоким сроком просроченной задолженности в разрезе рейтинговых групп.



Таблица 22. Частота переходов в разрезе рейтинговых групп

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	Выход на просрочку	Переход в группу 31-60 дней	Переход в группу 61-90 дней	Переход в группу свыше 90 дней
1	0	1,1%	21,0%	66,1%	77,6%
2	1	1,5%	25,3%	66,3%	77,1%
3	2	1,9%	30,5%	70,8%	82,3%
4	3	2,4%	31,8%	69,1%	78,4%
5	4	2,9%	36,6%	73,2%	81,6%
6	Более 5	4,6%	43,4%	77,3%	78,4%

Из таблицы 22 видно, что с увеличением числа членов комитета «проголосовавших» за дефолт увеличивается частота выхода на просрочку и частота перехода в группу 31-60 дней, при этом частота перехода в группу свыше 90 дней практически не зависит рейтинговой группы.

Далее рассчитаем вероятности дефолта исходя из частот перехода из одной группы просроченной задолженности в другую в разрезе рейтинговых групп и сравним с фактической долей дефолтов в каждой рейтинговой группе.

Таблица 23. Сравнение расчетного и фактического уровня дефолтов в разрезе рейтинговых групп

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	Расчетная вероятность дефолта	Фактическая доля заемщиков в дефолте
1	0	0,95%	0,76%
2	1	1,53%	1,63%
3	2	2,60%	2,51%
4	3	3,14%	3,47%
5	4	4,79%	6,16%
6	Более 5	8,85%	11,50%

Согласно таблице 23 расчетная и фактическая вероятности дефолта близки со второй по четвертую рейтинговые группы, при этом в пятой и шестой рейтинговых группах расчетная вероятность дефолта ниже, чем фактическая. Одной из возможных причин данного расхождения является то, что у нас

имеются данные о параметрах 72 052 кредитных договоров, выданных в период с 01.02.2011 по 01.08.2012 при этом, для расчета матриц миграций, учитываются данные о состоянии указанных кредитных договоров в период с 01.02.2012 по 01.08.2013. Заметим, что при расчете фактической доле дефолта нами использовалась информация о наличии или отсутствии просроченной задолженности по кредиту свыше 90 дней в течении 12 месяцев с момента выдачи, при этом при расчете матриц миграций нами рассматривались данные о частоте переходов кредитов в период превышающий 12 месяцев.

С целью минимизации влияния различий в периодах анализа попробуем построить матрицы миграций для кредитов, выданных в период с 01.01.2012 по 01.08.2012, в течение периода в 12 месяцев с момента выдачи. В таблице 24 приведено сравнение расчетной и фактической вероятностей дефолта по 37 119 кредитным договорам, выданным в период с 01.01.2012 по 01.08.2012, в разрезе рейтинговых групп.

Таблица 24. Сравнение расчетного и фактического уровня дефолтов в разрезе рейтинговых групп

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	Расчетная вероятность дефолта	Фактическая доля заемщиков в дефолте
1	0	0,95%	1,04%
2	1	1,96%	2,23%
3	2	2,92%	3,08%
4	3	3,63%	4,25%
5	4	5,65%	6,37%
6	Более 5	8,73%	11,53%

Как видно из таблицы 24 корректировка временных периодов не привела к значительному изменению ни в расчетной вероятности дефолта, ни в фактической вероятности, за исключением увеличения расчетной вероятности дефолта в четвертой рейтинговой группе и роста фактической вероятности дефолта в 1 рейтинговой группе.

Следовательно, выдвинутое ранее предположение о причинах расхождений в различиях временных периодов было некорректным. Вероятнее всего расхождение в основном объясняется тем, что при моделировании большее влияние оказывают кредиты, не перешедшие в дефолт, поскольку кредитные договоры после перехода в дефолт не учитываются моделью, в связи с чем в группах с высоким уровнем дефолта расчетная вероятность дефолта ниже, чем фактическая.

Исходя из результатов сравнения фактической вероятности дефолта и расчетной вероятности дефолта по матрицам миграций можно сделать вывод, что использование матриц миграций для уточнения рейтинговой модели не допустимо, поскольку при использовании матриц миграций вероятность дефолта снижается по сравнению с фактической, что связано с особенностью применяемой динамической модели.

### *3.2.5. Итоговая модель оценки кредитоспособности физического лица*

В данной части диссертационной работы будет описано построение рейтинговой модели, которая в дальнейшем нашла применение для решения практических задач оценки кредитного риска заемщика физического лица.

Как видно из трехмерного рисунка 30, приведенного выше в части 3.1.1., в пространстве 3 компонент не составляет сложности выделить группу заемщиков, в которой доля дефолтных будет минимальна, тогда в дальнейшем данная группа может быть исключена. Для решения данной задачи будет применен комитет состоящий из 1 члена, при этом для корректного построения члена, отделяющего только точки из недефолтного класса необходимо провести преобразования, а именно изменить параметры целевой функции, в задаче линейного программирования. Поскольку мы собираемся выделить множество кредитов, к которых доля дефолтных заемщиков будет минимальна, то необходимо увеличить штраф за некорректную классификацию дефолтных заемщиков. В данном случае целевая функция, выглядит следующим образом:

$$\min \left( \sum_j v_j * R_1 * E + \sum_i v'_i * R_2 \right) \quad (36)$$

где  $E=2$ ,  $R_1, R_2$  – число наблюдений в дефолте и без дефолта соответственно. Введение  $E=2$  говорит, о том, что допускается построение гиперплоскости, в которой вес ошибки по дефолтным наблюдениям, с учетом веса наблюдения, будет 2, при этом вес ошибки по недефолтным наблюдениям будет 1.

В таблице 25 приведены результаты тестирования комитетного решения на выборке из 62 614, из них дефолтных 1200.

Таблица 25. Качество решения одним членом комитета.

Голосование комитета	Наблюдения со случаями дефолта	Наблюдения без случаев дефолта	Доля заемщиков в дефолте
«Против» дефолта	57 288	907	1,6%
«За» дефолт	4 126	293	6,6%
Итого	61 414	1 200	1,9%

Как видно из таблицы 25, построенный комитет из одного члена позволяет разбить общее множество заемщиков на две группы, исходя из того, проголосовал член комитета «за» дефолт или «против» дефолта. В группе, по которой член комитета проголосовал за дефолт, доля заемщиков в дефолте более чем в 4 раза выше, чем в другой группе. Заметим, что одному члену комитета не удалось выявить группу с достаточно низким уровнем дефолтов или группу с достаточно высоким уровнем дефолта, следовательно, не представляется целесообразным исключать какую-либо из групп.

Построим член комитета, в котором целевая функция будет определена следующим образом:

$$\min \sum_j v_j * R_1 + \sum_i v'_i * R_2 * E \quad (37)$$

То есть, больший вес ошибки будет приходиться на заемщиков без дефолтов. Рассмотрим совместное голосование 2 членов комитета, в таблице приведены результаты голосования на тестовой выборке.

Таблица 26. Качество решения 2 членами комитета.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
0	1,3%	86,0%
1	5,0%	11,3%
2	9,3%	2,7%
Итого:	1,9%	100%

Из таблицы 26 видно, что в группе заемщиков, по которым 2 члена комитета проголосовали за дефолт, доля заемщиков в дефолте составила 9,3%. Попробуем увеличить качество комитетного решения за счет методов последовательной оптимизации, описанных ранее (веса ошибок в данном случае для дефолтных и недефолтных множеств равны).

Таблица 27. Качество решения после последовательной оптимизации.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
0	1,1%	75,0%
1	4,0%	23,3%
2	11,4%	1,6%
Итого:	1,9%	100%

Последовательная оптимизация комитетного решения позволила выделить группу заемщиков в которой доля дефолтных заемщиков составила 11,4%. Проведем сравнение данного комитетного решения с комитетным решением, полученным при построении р-комитета из 2 членов.

Таблица 28. Качество решения 2 членов комитета.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке

0	1,0%	71,4%
1	3,6%	22,5%
2	7,0%	6,1%
Итого:	1,9%	100%

Комитетное решение, построенное при обучении 2 членов комитета одновременно, обладает лучшим качеством классификации по сравнению с комитетным решением, полученным при последовательном построении 2 членов, поскольку позволяет точнее выделить заемщиков с уровнем дефолта почти в 2 раза меньше чем в среднем по тестовой выборке (1,0% против 1,9%).

Поскольку нами была выделена группа заемщиков с низким уровнем дефолта (1,0%) целесообразно рассматривать данную группу отдельно от остальных заемщиков (далее множество 1). Разбиение общего множества на 2 подмножества значительно уменьшить время, требуемое для построения комитета, поскольку необходимо будет решать задачи меньшей размерности.

Интересным будет рассмотреть, как распределены в пространстве 3 компонент множество заемщиков (далее множество 1), в которых за дефолт проголосовало 0 членов комитета, и множество заемщиков за дефолт которых проголосовал хотя бы 1 член комитета (далее множество 2)

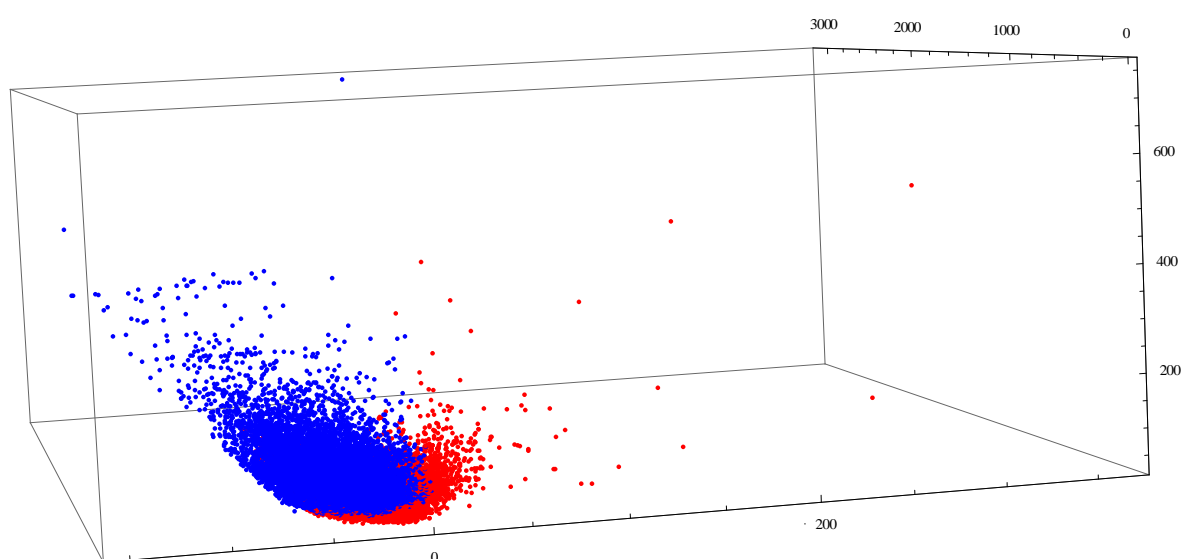


Рис. 36. Распределение заёмщиков в пространстве 3 компонент.

Синими точками на рисунке 36 обозначены заемщики из множества 1, а красными – заемщики из множества 2. Как видно из рисунка 34, заемщики, за дефолт которых проголосовало 0 членов комитета, представляют собой некоторое выпуклое множество.

### 3.2.5.1. Построение комитетного решения на множестве 1

Далее перейдем к построению комитетного решения на множестве 1 (заемщики с низким уровнем дефолта). Всего во множестве 1 содержится 44 722 наблюдений, из них дефолтных – 429. Обучающая выборка включает в себя 1000 наблюдений без случаев дефолта и 100 наблюдений со случаями дефолта. Последовательно построим комитеты из 3, 5 и 7 членов (построение комитета из большего числа членов признано было нецелесообразным, в связи с увеличением времени расчетов и дополнительной сложностью в интерпретации модели), для построения каждого из комитетов для нахождения первоначального решения используется генетический алгоритм, также для каждого из комитетов проводится последовательная оптимизация комитетного решения, и сравним качество полученных решений.

Таблица 29. Качество решения комитетом из 3 членов на множестве 1.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
0	0,70%	67,21%
1	1,40%	30,21%
2	2,52%	2,57%
Итого:	<b>0,96%</b>	<b>100,00%</b>

Таблица 30. Качество решения комитетом из 5 членов на множестве 1.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
0	0,68%	66,92%
1	2,03%	8,28%
2	1,18%	15,53%

3	1,26%	6,73%
4	2,46%	2,54%
Итого:	<b>0,96%</b>	<b>100,00%</b>

Таблица 31. Качество решения комитетом из 7 членов на множестве 1.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
0	0,47%	47,00%
1	0,85%	27,87%
2	1,68%	18,12%
3	2,71%	2,84%
4	3,10%	2,65%
5	2,03%	1,5%
Итого:	<b>0,96%</b>	<b>100,00%</b>

Как следует из таблиц 29, 30 и 31 комитет из 7 членов имеет наибольшую дискриминантную силу, в дальнейшем данный комитет будет использован для построения рейтинговой модели оценки кредитоспособности заемщиков. Комитеты из большего числа членов, использовать не целесообразно в связи со значительным увеличением времени расчетов.

Исходя из полученного решения комитета из 7 членов, можно построить следующую рейтинговую модель (напомним, что PD группы соответствует среднему уровню дефолтов в группе):

Таблица 32. Рейтинговая модель для множества 1.

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	PD группы	Доля группы в тестовой выборке
1	0	0,51%	32,00%
2	1	0,76%	27,87%
3	2	1,39%	36,60%
4	3 и более членов	2,03%	3,53%
Итого:		0,96%	100,00%



### 3.2.5.2. Построение комитетного решения на множестве 2

Далее перейдем к построению рейтинговой модели для множества 2 (заемщики с высоким уровнем дефолтов). Всего во множестве 1 содержится 21 969 наблюдений, из них дефолтных – 558. Обучающая выборка включает в себя 1000 наблюдений без случаев дефолта и 100 наблюдений со случаями дефолта. Последовательно построим комитеты из 3, 5 и 7 членов, для построения каждого из комитетов для нахождения первоначального решения используется генетический алгоритм, также для каждого из комитетов проводится последовательная оптимизация комитетного решения, и сравним качество полученных решений.

Таблица 33. Качество решения комитетом из 3 членов на множестве 1.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
0	2,5%	17,64%
1	4,6%	61,48%
2	7,2%	20,88%
Итого:	<b>4,31%</b>	<b>100,00%</b>

Таблица 34. Качество решения комитетом из 5 членов на множестве 1.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
1	3,5%	60,57%
2	6,9%	5,84%
3	6,7%	23,87%
4	5,8%	9,73%
Итого:	<b>4,31%</b>	<b>100,00%</b>

Таблица 35. Качество решения комитетом из 7 членов на множестве 1.

Число членов комитета «проголосовавших» за дефолт	Доля заемщиков в дефолте	Доля группы в тестовой выборке
0	1,9%	4,67%

1	3,4%	22,57%
2	4,4%	46,82%
3	6,5%	7,00%
4	6,0%	12,19%
5	12,7%	6,74%
Итого:	<b>4,31%</b>	<b>100,00%</b>

Как следует из таблиц 33, 34 и 35 комитет из 7 членов имеет наибольшую дискриминантную силу, в дальнейшем данный комитет будет использован для построения рейтинговой модели оценки кредитоспособности заемщиков.

Таблица 36. Рейтинговая модель для множества 2.

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	PD группы	Доля группы в тестовой выборке
1	0	1,91%	10,76%
2	1	3,44%	28,91%
3	2	4,43%	46,59%
4	3-4	6,16%	13,73%
5	5 и более	12,75%	2,33%
Итого:		<b>4,31%</b>	<b>100,00%</b>

После построения рейтинговых моделей для множества 1 (заемщики с низким уровнем дефолта) и множества 2 (заемщики с высоким уровнем дефолта) необходимо составить единую рейтинговую модель оценки вероятности дефолта заемщика. Напомним, что в ходе работы было построено 3 комитета – комитет, определяющий принадлежность к множествам 1 или 2 (далее комитет 1), и 2 комитета для оценки вероятности дефолта в обоих множествах (комитеты 2 и 3 соответственно).

Таблица 37. Итоговая рейтинговая модель.

Рейтинговая группа	PD группы	Доля группы в тестовой выборке
1	0,51%	22,85%
2	0,76%	19,90%

3	1,39%	26,14%
4	1,96%	5,53%
5	3,44%	8,08%
6	4,43%	13,02%
7	6,16%	3,84%
8	12,75%	0,65%
<b>Итого:</b>	<b>1,91%</b>	<b>100,00%</b>

Для наглядности итоговая рейтинговая модель представлена в виде некоторого древа принятия решений.

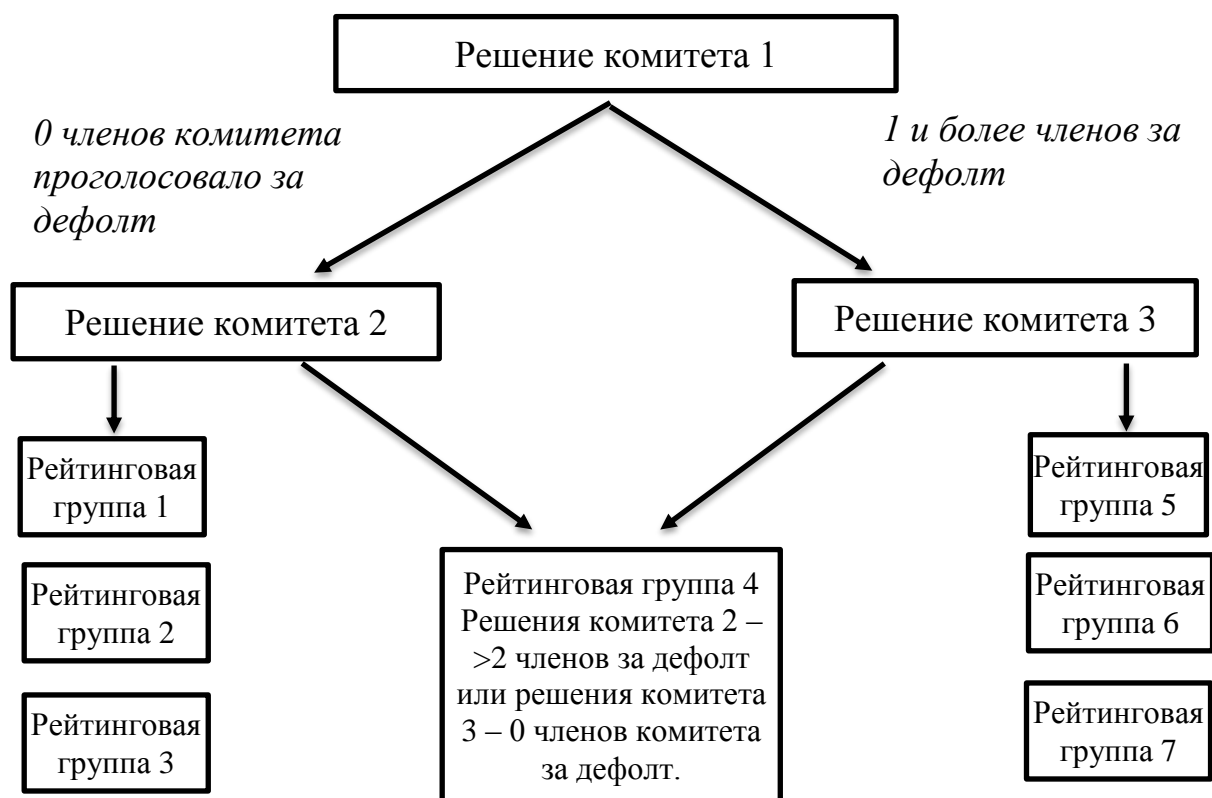


Рис. 37. Рейтинговая модель.

Для сравнения приведем параметры рейтинговой модели, полученной исходя из решений комитета из 7 членов, в случае если множество заемщиков предварительно не разбивалось на 2 подмножества.

Таблица 38. Рейтинговая модель, построенная исходя из решения комитета из 7 членов

Рейтинговая группа	Число членов комитета «проголосовавших» за дефолт	PD группы	Доля группы в тестовой выборке
1	0	0,95%	64,50%
2	1	1,79%	4,30%
3	2-3	2,53%	6,50%
4	4	4,42%	15,40%
5	5 и более	6,32%	9,30%
Итого:		1,91%	100,00%

Исходя из сравнения таблиц 37 и 38 видно, что разбиение исходного множества на 2 подмножества позволило построить модель с большей дискриминационной силой.

### 3.3. Валидация рейтинговых моделей.

В данной главе будет описан процесс валидации рейтинговой модели, полученной ранее во второй главе. Под термином «валидация» понимается проверка соответствия данных, получаемых в процессе машинной имитации, реальному ходу явлений, для описания которых создана модель. Состоит в том, что выходные данные после расчета на компьютере сопоставляются с имеющимися статистическими сведениями о моделируемой системе.

Исходя из представленного выше определения валидации как проверки соответствия модели реальному ходу событий, цель валидации рейтинговой модели заключается в подтверждении соответствия рейтинговой системы нормативным требованиям и адекватности вычисляемых риск – параметров.

Группой российских экспертов банковской отрасли и представителей ведущих российских коммерческих банков в составе Комитета по стандартам Базель II и управлению рисками при Ассоциации российских банков разработаны требования к валидации рейтинговых моделей [99]. Основной целью деятельности указанной экспертной группы является выработка рекомендаций, позволяющих эффективно внедрить статистические модели, которые соответствуют стандартам Базель II и Письму Банка России от 29.12.2012 № 192-

Т "О Методических рекомендациях по реализации подхода к расчету кредитного риска на основе внутренних рейтингов банков".

На основании рекомендаций Базельского Комитета были сформулированы принципы проведения валидации [2,99]:

1. В первую очередь банк является ответственным за проведение валидации;
2. Валидация, по своей сути, является оценкой предсказательной способности риск-параметров, рассчитанных банком, а также их применения в кредитном процессе;
3. Валидация – итеративный процесс;
4. Не существует единого подхода по проведению валидации;
5. Валидация должна включать как количественную, так и качественную оценку;
6. Следует проводить независимую проверку процесса и результатов валидации.

Процесс валидации рейтинговой модели подразумевает под собой как количественную, так и качественную оценку модели.

Качественная оценка модели включает в себя оценку следующих составляющих модели: документацию модели, методологию разработки модели, определение дефолта, глубину и качество данных, использованных для разработки, подходы к калибровке, применяемая рейтинговая шкала, применяемые корректировки рассчитанного рейтинга.

Количественная валидация моделей включает в себя оценку следующих составляющих модели: дискриминационная способность модели, калибровка модели, репрезентативность, концентрация, стабильность, влияние процесса на результаты работы модели.

Различают 3 типа валидации [99]:

1. Первоначальная - валидация, проводимая при разработке/до ввода в действие модели, а также в случае существенных<sup>2</sup> изменений модели. Данный тип валидации включает в себя тесты по структуре модели и

тесты на эффективность модели, оценивает ее соответствие, по меньшей мере, минимальным требованиям.

2. Мониторинг - с частой периодичностью проводимая валидация (например, ежемесячная, ежеквартальная), включающая в себя проведение тестов высокого уровня и направленная на выявление ранних признаков ухудшения эффективности модели.
3. Последующая - ежегодная углубленная валидация, включающая проведение всех групп тестов для подтверждения адекватного функционирования модели в случае отсутствия существенных изменений в модели. Разработчики модели вправе запросить внеплановую последующую валидацию в случае выявления признаков значительного ухудшения модели в рамках процесса мониторинга.

Далее будут проведены основные тесты для количественной и качественной оценки качества рейтинговой модели, построенной с использованием метода комитетов.

Для валидации рейтинговой модели использовались следующие показатели: скорректированный индекс Херфиндаля, меры AUROC, AR, IV. Так значение скорректированного индекса Херфиндаля (индекса концентрации) составило 7,2% (при допустимом значении для принятия рейтинговой модели – не более 30% [99]) значение мер дискриминационной силы AUROC и AR составило 72,3% и 45,9% соответственно (допустимые значения – более 70% и 40% соответственно [100]), значение меры IV составило 274% (допустимые значения не установлены).

В целом рейтинговая модель, построенная с использованием метода комитетов, удовлетворяет основным условиям для принятия ее к использованию в рамках IRB подхода.

### *3.3.1. Индекс Херфиндаля*

Одним из тестов, позволяющих количественно оценить качество рейтинговой модели, расчет индекса Херфиндаля. Индекс Херфиндаля позволяет

выявить избыточную концентрацию кредитов на некоторых интервалах рейтинговой шкалы.

Индекс Херфиндаля рассчитывается по следующей формуле [99]:

$$HI = \sum_i \left( \frac{N_i}{N} \right)^2 \quad (37)$$

Где  $i$  – индекс по интервалам рейтинговой шкалы,  $N_i$  – число наблюдений в  $i$ -том интервале,  $N$  – общее число наблюдений.

Исходя из данных, представленных в таблице 38, рассчитаем индекс Херфиндаля. Значение индекса составило 18,8%.

Одной из проблем применения индекса Херфиндаля является то, что в индексе не учитывается число интервалов в рейтинговой модели, в связи с чем увеличение числа рейтингов приводит к уменьшению индекса Херфиндаля. Обычно для оценки концентрации кредитов используют скорректированный индекс Херфиндаля, рассчитываемый по следующей формуле [99]:

$$HI^{adj} = \frac{\sum_i \left( \frac{N_i}{N} \right)^2 - \frac{1}{j}}{1 - \frac{1}{j}} \quad (38)$$

Где  $j$  – число интервалов рейтинговой шкалы.

Значение скорректированного индекса Херфиндаля составило 7,2%. Пороговым значением индекса Херфиндаля для принятия рейтинговой в рамках IRB подхода считается 30% [99] То есть, полученная нами модель в целом удовлетворяет условиям по индексу Херфиндаля.

### 3.3.2. AUROC

AUROC (Area Under the Receiver Operating Characteristic) – площадь под ROC кривой является показателем, характеризующим качество ранжирования в модели. ROC кривая определяется следующим образом: все наблюдения в дефолте и не в дефолте упорядочены в соответствии с рейтинговым баллом от наихудшего к наилучшему, тогда точки ROC кривой определяются как

соотношения доли недефолтных заемщиков к дефолтным [99]. На рисунке 38 представлен график ROC кривой для полученной рейтинговой модели.

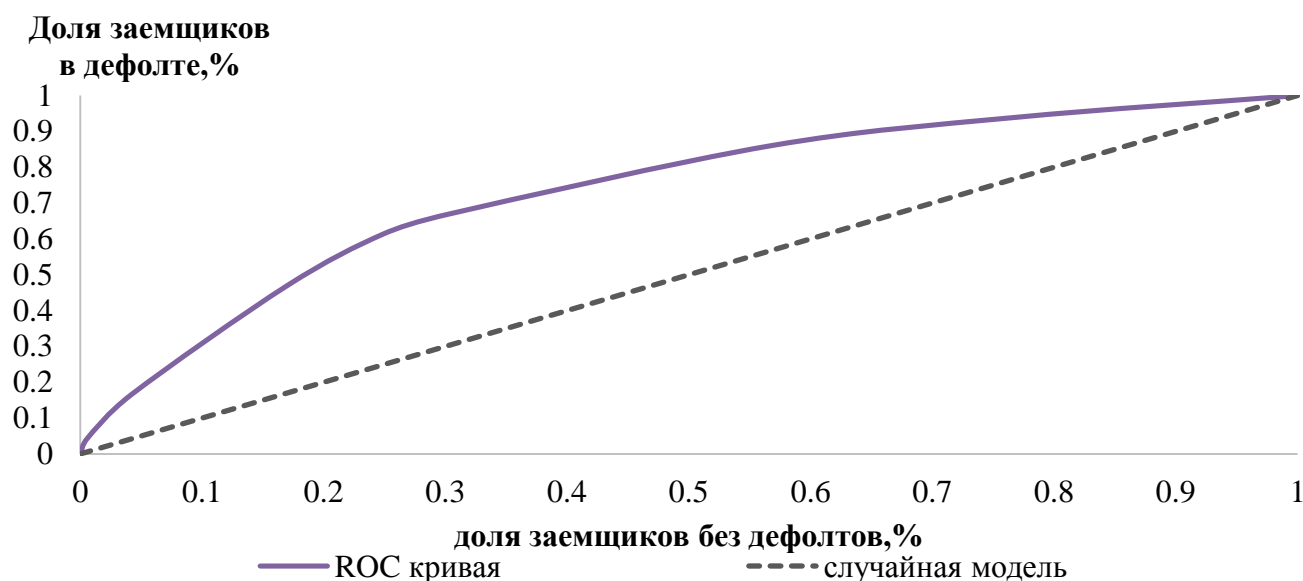


Рис. 38. ROC кривая

Площадь под ROC кривой позволяет определить на сколько построенная модель отличается от случайной. Значения показателя AUROC варьируется от 0 (модель работает с точностью наоборот) до 1 (модель работает идеально).

Значение показателя AUROC вычисляется по следующей формуле [99]:

$$A = \frac{\sum_{j,i} B_i * G_j * a_{ij}}{B * G},$$

$$a_{ij} = \begin{cases} 1, & \text{если } i < j \\ 0.5, & \text{если } i = j \\ 0, & \text{если } i > j \end{cases} \quad (39)$$

Где  $B_i$ ,  $G_j$  – число наблюдений в дефолте и без дефолта для  $i$ -го и  $j$ -го аргумента соответственно.  $B$  и  $G$  – общее число наблюдений в дефолте и без дефолта.  $i$ -го и  $j$ -го аргументы являются порядковыми номерами рейтинговых групп, отсортированных по возрастанию вероятности дефолта. Рассчитанное значение показателя AUROC равняется 72,3%.

Далее проведем расчет стандартной ошибки показателя AUROC, для чего воспользуемся следующей формулой:



$$\sigma_{AUROC} = \sqrt{\frac{A(1-A) + (B-1)\left(\frac{A}{2-A} - A^2\right) + (G-1)\left(\frac{2A^2}{1+A} - A^2\right)}{BG}} \quad (40)$$

Значение стандартной ошибки показателя AUROC составило 0,8%, что является допустимым.

Часто вместо показателя AUROC используется показатель AR или индекс Джини, который так же характеризует качество модели, ранее данный показатель был использован во второй главе для отбора моделей. Показатель AR рассчитывается как  $AUROC * 2 - 1$ . Значения показателя AR варьируется от -1 (модель работает с точностью наоборот) до 1 (модель работает идеально). Значение показателя AR для рейтинговой модели, построенной с применением метода комитетов равняется 44,6%.

Согласно аналитическому документу о степени соответствия внутрибанковских подходов к управлению кредитным риском банков – участников проекта «Банковское регулирование и надзор (Базель II)» Программы сотрудничества Евросистемы с Банком России минимальным требованиям IRB-подхода Базеля II [100] допустимыми значениями меры AUROC являются значения выше 70%, допустимыми значениями меры AR являются значения выше 40%.

### 3.3.3. Мера Information value (IV)

Еще одной из мер оценки качества рейтинговой модели является мера Information value (IV). Значение показателя IV рассчитывается по следующей формуле [94]:

$$IV = \sum_i (B\%_i - G\%_i) \ln\left(\frac{B\%_i}{G\%_i}\right) \quad (41)$$

Где  $B\%_i$  и  $G\%_i$  - доля дефолтных и недефолтных наблюдений от общего числа дефолтных и недефолтных наблюдений в  $i$ -том рейтинге соответственно.

Значение меры IV для построенной рейтинговой модели равняется 274%. Заметим, что граничных значений для данной меры не существует. IV может принимать значения от 0 до бесконечности, при этом низкие значения соответствуют низкой различительной способности модели. Достоинством этой меры является простота расчета и возможность оценки различительной способности для каждого фактора, но при этом оценка является смещенной.

#### *3.3.4. Сравнение рейтинговой модели основанной на комитетных конструкциях с рейтинговой моделью, построенной с применением Logit метода*

В данной части работы будет проведено сравнение полученной рейтинговой модели, с рейтинговой моделью, построенной с использованием Logit метода. Сравнение с Logit методом производится связи с частотой его использования для построения рейтинговых моделей, в частности в ПАО «Сбербанк» используется рейтинговая модель, основанная на Logit методе.

Для Logit метода сформируем обучающую выборку, составляющую около 40% от генеральной совокупности (28 000 заемщиков без случаев дефолта и 560 заемщиков со случаями дефолта). Заметим, что доля заемщиков в дефолте в обучающей выборке составляет 2%. Для применения Logit метода требуется, чтобы доля каждого класса превышала 30%, для этого искусственно увеличим долю заемщиков в дефолте за счет дублирования дефолтных наблюдений. Так после дублирования дефолтных наблюдений обучающая выборка составила (28 000 заемщиков без случаев дефолта и 28 000 заемщиков со случаями дефолта).

Приведем результаты Logit метода на тестовой выборке.

Таблица 39. Результаты рейтинговой модели, построенной с использованием Logit метода

Рейтинговая группа	Вероятность по Logit методу	PD группы	Доля группы в тестовой выборке
1	от 0% до 20%	0,6%	5,3%
2	от 20% до 30%	0,7%	19,9%

3	от 30% до 40%	0,9%	20,8%
4	от 40% до 50%	1,6%	17,0%
5	от 50% до 60%	2,4%	16,7%
6	от 60% до 70%	4,2%	14,3%
7	от 70% до 100%	8,5%	6,0%
Итого:		2,1%	100,0%

В таблице ниже приведены значения валидационных тестов по рейтинговой модели, построенной с применением Logit метода и рейтинговой модели, построенной с применением комитетных конструкций.

Таблица 40. Сравнение рейтинговых моделей

Показатель	Logit метод	Комитетные конструкции
Индекс херфиндаля	2,80%	7,20%
AUROC	69,1%	72,3%
AR	38,1%	45,9%
IV	68,0%	274,0%

Как видно из таблицы 40, значения AUROC, AR, IV рейтинговой модели, построенной с применением комитетных конструкций выше, чем у рейтинговой модели, построенной с применением Logit метода, что говорит о большом качестве модели. Уровень концентрации у модели, построенной с применением Logit метода меньше, чем у модели, построенной с применением метода комитетов.

### 3.4. Оценка экономического эффекта от внедрения рейтинговой модели

Важным для внедрения рейтинговой модели оценки заемщиков является расчет экономического эффекта от внедрения. В данной части работы будет произведен расчет экономического эффекта от внедрения предложенной рейтинговой модели оценки кредитного риска заемщика.

### 3.4.1. Расчет взвешенных по риску активов и капитала

Как уже было отмечено ранее в главе 1 Банком России осуществляется внедрение базельских подходов к оценке кредитных рисков и капитала, в связи с чем рассмотрим экономический эффект от снижения требований к капиталу кредитной организации. Расчет величины взвешенных по риску кредитных требований осуществляется с помощью следующей формулы:

$$RWA = 12,5 * EAD * LGD * \left( N \left( \frac{(N^{-1}(PD) + \sqrt{R} * N^{-1}(0,999))}{\sqrt{1-R}} \right) - PD \right) \quad (42)$$

где R - показатель корреляции, значение которого установлено равным:

0,04 - для активов, отнесенных к подклассу "возобновляемые розничные кредитные требования"; 0,15 - для активов, отнесенных к подклассу "кредитные требования, обеспеченные залогом жилой недвижимости".

Значение показателя корреляции для кредитных требований, отнесенных к подклассу "прочие розничные кредитные требования", рассчитывается по следующей формуле:

$$R(PD) = 0,03 * \frac{1 - e^{-35*PD}}{1 - e^{-35}} + 0,16 * \left( 1 - \frac{1 - e^{-35*PD}}{1 - e^{-35}} \right) \quad (43)$$

Требования к капиталу соответственно рассчитываются по следующей формуле:

$$K = LGD * \left( N \left( \frac{(N^{-1}(PD) + \sqrt{R} * N^{-1}(0,999))}{\sqrt{1-R}} \right) - PD \right) \quad (44)$$

Расчет суммы высвобождаемого капитала производится по следующей формуле:

$$\Delta Capital = EAD * K_1 - EAD * K_2 \quad (45)$$

где K1 и K2 требования к капиталу, рассчитанные до и после применения рейтинговой модели соответственно.

Оценку экономического эффекта от высвобождаемого капитала можно провести исходя из средней стоимости собственного капитала или исходя из увеличения кредитного портфеля без увеличения собственного капитала.

Экономический эффект исходя из средней стоимости капитала рассчитывается по следующей формуле:

$$Effect_{cap} = \Delta Capital * Cap_{cost}, \% \quad (46)$$

где  $Cap_{cost}, \%$  – средняя стоимость собственного капитала банка.

Заметим, что корректный расчет средней стоимости собственного капитала является затруднительным, в связи с чем оценку экономического эффекта целесообразнее проводить исходя из оценки увеличения кредитного портфеля без увеличения собственного капитала.

Экономический эффект исходя из увеличения кредитного портфеля без увеличения собственного капитала рассчитывается по следующей формуле:

$$Effect_{EAD} = EAD * \frac{RWA_1}{RWA_2} * (Nim - PD - OPEX) \quad (47)$$

где  $RWA_1$  и  $RWA_2$  – взвешенные по риску активы, рассчитанные до и после применения рейтинговой модели соответственно,  $Nim$  – чистая процентная маржа,  $PD$  – вероятность дефолта по портфелю в целом,  $OPEX$  – операционные расходы.

В таблице 41 приведены значения показателей, учитываемых при расчете взвешенных по риску активов.

Таблица 41. Расчет взвешенных по риску активов.

Рейтинговая группа	PD	EAD (объем кредитного портфеля), тыс. руб.	RWA (взвешенные по риску активы)	Требования к капиталу
1	0,51%	3 676 027	2 004 886	4,4%
2	0,76%	3 471 764	2 335 266	5,4%
3	1,39%	4 663 711	4 043 309	6,9%
4	1,96%	553 069	531 794	7,7%
5	3,44%	1 326 627	1 413 907	8,5%
6	4,43%	2 137 867	2 339 513	8,8%
7	6,16%	630 003	713 551	9,1%
8	12,75%	107 011	147 446	11,0%
Итого по	1,91%	16 566 079	13 529 672	6,5%

рейтинговой модели:				
Итого в случае отсутствия рейтинговой модели	1,91%	16 566 079	15 822 847	7,6%

Уровень потерь в случае дефолта для всех рейтинговых групп принят равному 75%.

Из таблицы 41 видно, что взвешенные по риску активы, в случае расчета по всему портфелю в целом, составили 15 822 847 тыс. руб. (требования к капиталу – 7,6%), тогда как взвешенные по риску активы, рассчитанные по рейтинговой модели, составляют 13 529 672 тыс. руб. (требования к капиталу – 6,5%). Следовательно, внедрение рейтинговой модели позволило бы кредитной организации снизить взвешенные по риску активы с 16 566 079 до 13 529 672 (- 2 293 176 тыс. руб. / -14.5%), что позволило бы увеличить объем кредитного портфеля на 14,5% (с 16 566 079 тыс. руб. до 18 966 970).

Учитывая, что средняя ставка по выданным кредитам в период с 01.01.2011 по 01.01.2013 составляла 21,1% годовых, средняя стоимость привлечения кредитных ресурсов - 11,9% годовых, операционные расходы на выдачу потребительских кредитов (с учетом распределения управленческих расходов) составили 298 189 тыс. руб. (1,8% от объема выданных кредитов), то маржа кредитного портфеля составляла 7,4% (21,1% – 11,9% – 1,8%), маржа с учетом кредитного риска (PD) равнялась 5,5%. Тогда дополнительная прибыль от увеличения кредитного портфеля на 2 400 891 тыс. руб. составила бы 132 049 тыс. руб.

#### *3.4.2. Расчет экономического эффекта от отказа в кредитовании заемщиков с высоким уровнем риска*

Проведем расчеты экономического эффекта от отказа в кредитовании заемщиков с высоким уровнем кредитного риска, для этого сравним маржу с учетом риска в каждой из рейтинговых групп. Как уже было указано выше маржа

без учета риска составила в среднем 7,4%. В таблице 42 приведены значения маржи с учетом риска и объемы прибыли / убытков в разрезе рейтинговых групп.

Таблица 42. Расчет прибыли / убытков по рейтинговым группам.

Рейтинговая группа	PD, %	Объем кредитного портфеля, тыс. руб.	Маржа с учетом риска, %	Объем прибыли / убытков, тыс. руб.
1	0,51%	3 676 027	6,89%	253 278
2	0,76%	3 471 764	6,64%	230 525
3	1,39%	4 663 711	6,01%	280 289
4	1,96%	553 069	5,44%	30 087
5	3,44%	1 326 627	3,96%	52 534
6	4,43%	2 137 867	2,97%	63 495
7	6,16%	630 003	1,24%	7 812
8	12,75%	107 011	-5,35%	-5 725

Исходя из таблицы 42, только одна рейтинговая группа (наиболее рискованные заемщики), имеет уровень риска, превышающий маржу, в связи с чем по данной группе наблюдаются убытки в размере -5 725 тыс. руб., то есть отказ от кредитования данной группы увеличил бы прибыль кредитной организации. Заметим также, что на первые 3 группы заемщиков приходится 83,8% от совокупной прибыли и 71,3% кредитного портфеля, тогда как на заемщиков остальных групп приходится только 16,2% совокупной прибыли и 28,7% кредитного портфеля. В целом формирование кредитной стратегии, направленной на кредитование только заемщиков первых 3 рейтинговых групп, положительно сказалось на деятельности кредитной организации и позволило бы увеличить маржу с учетом риска с 5,5% до 6,5%.

### 3.5. Алгоритмы и принципы применения рейтинговой модели

#### 3.5.1. Алгоритм построения рейтинговой модели

Построение рейтинговой модели с применением инструментария комитетных конструкций предлагается проводить по следующему алгоритму.

**Этап 1. Обработка исходных данных.** На этапе 1 осуществляется предварительная работа с имеющейся генеральной совокупности по кредитным договорам. Из общей совокупности необходимо исключить дефолтных заемщиков, допустивших просроченную задолженность с первого платежа, поскольку, по мнению автора, дефолт данных заемщиков связан вероятнее всего с мошенническими операциями, проверка заемщиков на предмет возможного мошенничества должна осуществляться на этапе «фрод-мониторинга». Учет данных заемщиков в модели может исказить общую картину. Также на этапе обработки данных необходимо исключить заемщиков с ошибками/отклонениями в данных, так очевидно, что необходимо исключить заемщиков в возрасте свыше 1000 лет. Исходя из экономического смысла добавляются дополнительные переменные, такие как отношение доходов к кредиту, отношение возраста к доходам, отношение месячного платежа к доходам.

**Этап 2. Выявление корреляций объясняющих переменных, уменьшение числа переменных.** На данном этапе необходимо проверить наличие корреляций между переменными, в случае если существует корреляция свыше 40%, целесообразно применить метод главных компонент с целью уменьшения числа переменных и снижения вычислительной сложности задачи.

**Этап 3. Построение рейтинговой модели.** На начальном этапе построения рейтинговой модели предлагается рассмотреть возможность разделения генеральной совокупности на 2 подмножества – заемщики с высоким уровнем риска и заемщики с низким уровнем риска. Для этого из генеральной совокупности формируется обучающая выборка, по которой строится комитет из одного до трех членов. Исходя из результатов комитетного решения на тестовой выборке, определяются множества заемщиков с высоким и низким уровнями риска. Далее на множестве с низким уровнем риска строится комитет из 7 членов, исходя из результатов комитетного решения на тестовой выборке, формируется рейтинговая модель для множества с низким уровнем риска. Аналогичные действия проводятся для множества заемщиков с высоким уровнем



риска. Полученные рейтинговые модели, группируются в единую рейтинговую модель, рейтинги распределяются по возрастанию уровня кредитного риска.

**Этап 4. Валидация рейтинговой модели.** Полученная рейтинговая модель проходит через процедуры валидации, в частности рассчитываются следующие показатели: скорректированный индекс Херфиндаля, меры AUROC, AR. В случае, если значения показателей соответствуют рекомендованным [94], рейтинговая модель может быть принята к использованию для оценки риска заемщиков.

### *3.5.2. Основные принципы дальнейшего применения рейтинговой модели*

Очевидно, что с течением времени, качество рейтинговой модели может снизиться, в связи с чем одним из условий применения рейтинговых моделей оценки кредитного риска является наличие принципов, позволяющих корректировать/перестраивать рейтинговую модель с течением времени. Также необходимым условием применения рейтинговой модели является корректировка уровней риска групп, исходя из текущего уровня рисков, зависящего от экономической ситуации. Исходя из рекомендаций [2,6,99,100] автором были разработаны основные принципы, позволяющие перестраивать модель с требуемой периодичностью.

Как было отмечено выше, применение матриц миграций позволяет оценить кредитные риски в целом по действующему портфелю. Подразделением, отвечающим за оценку кредитного риска, на ежемесячной основе производится оценка рисков кредитного портфеля на основе матриц миграций кредитов по группам просроченной задолженности. В случае отклонения уровня кредитного риска модели от расчетного по матрицам миграций более чем на 0,3 п.п., производится корректировка уровня рисков для каждой из групп пропорционально вкладу данной группы в совокупный риск портфеля, так чтобы кредитный риск, рассчитанный по рейтинговой модели, соответствовал текущему уровню риска.

Подразделением, отвечающим за разработку рейтинговых моделей на ежемесячной основе, производится корректировка весов модели, исходя из новой информации о случаях дефолтов. Корректировка производится по методу последовательной оптимизации, так для каждого из параметров членов комитета устанавливается допустимое отклонение от текущего значения и решается задача целочисленного программирования с ограничением допустимого изменения весов членов. Допустимое изменение параметров членов комитета устанавливается на уровне  $\pm 5\%$  от текущего уровня. Далее полученная рейтинговая модель проходит процедуру валидации (индекс Херфиндаля, меры AUROC и AR), и, в случае, если значение показателей AUROC и AR новой модели превосходит значения показателей AUROC и AR действующей модели осуществляется замена модели.

Раз в полугодие подразделением, отвечающим за разработку рейтинговых моделей, осуществляется построение новых комитетных конструкций и, соответственно, новой рейтинговой модели на основе новых данных о дефолтах. Период раз в полугодие выбран в связи с трудоемкостью построения модели, а также в связи с отсутствием практической целесообразности в частой смене кредитной политики банка. Замена рейтинговой модели должна производиться в случае, если качество новой рейтинговой модели (на новых данных по заемщикам) выше качества действующей рейтинговой модели (значение показателя AUROC выше на 3 п.п., AR на 6 п.п.).

Подразделением, отвечающим за валидацию рейтинговых моделей, на ежемесячной основе, исходя из новых данных о дефолтах заемщиков, производится оценка качества действующей модели. В случае, если отмечается, значительное снижение качества рейтинговой модели (снижение AUROC на 5 п.п., или AR на 10 п.п., превышение индекса Херфиндаля 30%) инициируется процедура построения новой рейтинговой модели.

### 3.6. Применение комитетных конструкций к принятию решений на финансовых рынках

В данной части работы рассмотрены возможности применения комитетных конструкций к принятию решений на финансовых рынках. Применим инструментарий, разработанный в рамках диссертационного исследования, для прогнозирования движения индекса CAC 40 (основной фондовый индекс Франции) на 10 минутном интервале. Поскольку метод комитетов предназначен в основном для решения задач классификации, то индекс CAC40 рассматривается как бинарная величина, которая имеет два возможных исхода: рост или падение. Для прогнозирования цены индекса в следующие 10 минут используются данные о динамике цены индекса за предшествующие 60 минут, что соответствует одной из аксиом технического анализа, а именно «движения цен учитывают всё». Заметим, что по мнению автора данную аксиому можно поставить под сомнение.

Имеются данные по цене закрытия по фондовому индексу CAC 40 в период с 17.04.2014 по 17.05.2015, общее число наблюдений в 10 минутном интервале составило 12 443. На обучающих выборках были применены методы технического анализа и метод комитетов для предсказания направления движения фондового индекса.

В качестве переменных, объясняющих динамику индекса, выбраны:

1.  $\frac{P(t-4)}{P(t-5)}$ , где  $P(t)$  – цена индекса в момент времени  $t$ . Шаг по  $t$  равен 10 минутам;
2.  $\frac{P(t-3)}{P(t-4)}$ ;
3.  $\frac{P(t-2)}{P(t-3)}$ ;
4.  $\frac{P(t-1)}{P(t-2)}$ ;
5.  $\frac{P(t_0)}{P(t-1)}$ ;

$$6. \frac{P(t_0)}{\min(P(t_{-1}), P(t_{-2}), \dots, P(t_{-6}))};$$

$$7. \frac{P(t_0)}{\max(P(t_{-1}), P(t_{-2}), \dots, P(t_{-6}))}$$

В качестве зависимой переменной рассматривается изменение цены за следующие 10 минут. Поскольку метод комитетов, предназначен для классификации 2 множеств, то зависимая переменная представлена бинарными значениями (1 – рост цены, 0 – снижение цены).

На обучающей выборке за выбран период с 17.04.2015 по 19.05.2014 (всего 1000 наблюдений на 10 минутном интервале) был построен комитет из 7 членов.

Таблица 43. Результаты комитетного решения на обучающей выборке.

Голосование комитета	Общее число наблюдений	Число верных решений	Число неверных решений	Доля правильных решений	Доверительный интервал (95%) точности
«За» рост	464	259	205	55,8%	±3,8%
«Против» роста	536	310	226	57,8%	±3,5%
<b>Итого:</b>	<b>1000</b>	<b>569</b>	<b>431</b>	<b>56,9%</b>	<b>±2,6%</b>

Как видно из таблицы 43 точность комитетного решения на обучающей выборке составила 56,9% (доверительный интервал точности от 54,3% до 59,5%), при этом точность в случае голосования за рост составила 55,8%, в случае голосования против роста – 57,8%, в целом комитет на обучающей выборке с одинаковой точностью прогнозирует как рост курса, так и снижение курса.

Полученная комитетная конструкция была применена для прогнозирования движения индекса на тестовой выборке за период с 19.05.2014 по 17.04.2015, содержащей 11 439 наблюдений. Периоды обучающей и тестовой выборок не пересекаются.

Результаты комитетного решения на тестовой выборке приведены в таблице 44.

Таблица 44. Результаты комитетного решения на тестовой выборке.

Голосование комитета	Общее число наблюдений	Число верных решений	Число неверных решений	Доля правильных решений	Доверительный интервал (95%) точности
«За» рост	5104	2565	2539	50,3%	$\pm 1,2\%$
«Против» роста	6335	3221	3114	50,8%	$\pm 1,0\%$
<b>Итого:</b>	<b>11439</b>	<b>5786</b>	<b>5653</b>	<b>50,6%</b>	<b><math>\pm 0,8\%</math></b>

Согласно таблице 44 точность комитетного решения на тестовой выборке снизилась до 50,6% (доверительный интервал от 49,8% до 51,4%), при том, что на обучающей выборке точность составляла 56,9%. Интересным будет рассмотреть изменение точности комитетного решения в течение рассматриваемого периода.

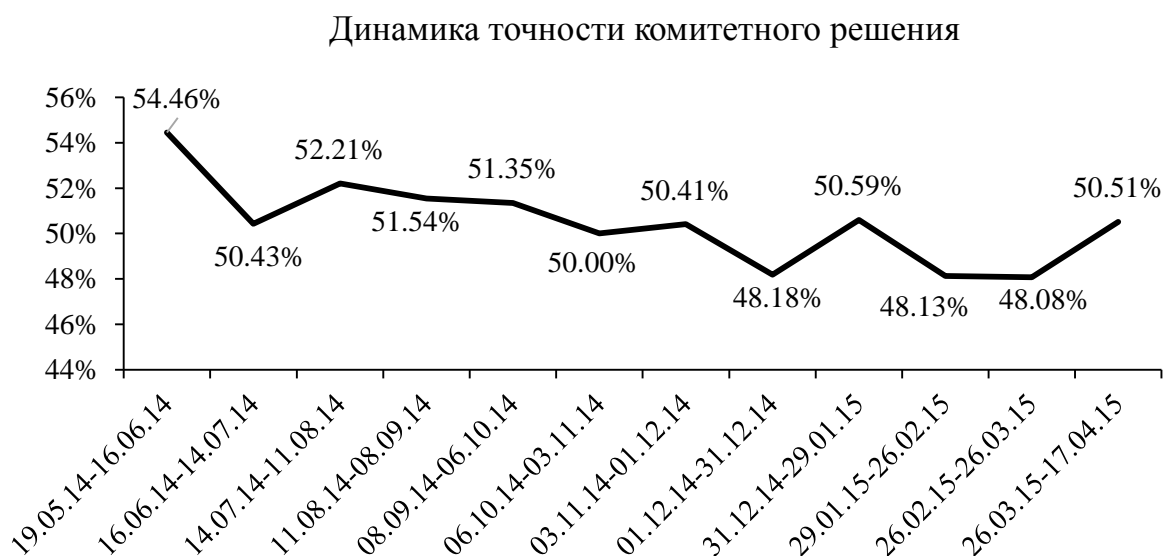


Рис. 39. Динамика точности комитетного решения.

Как видно из рисунка 39, точность комитетного решения имеет тенденцию к снижению в течение наблюдаемого периода. Максимальная точность (54,46%) комитетного решения на тестовой выборке достигается в первые 20 торговых дня (с 19.05.2014 по 16.06.2014). Точность комитетного решения на тестовой выборке снижается с течением времени, что может быть связано с изменением влияния объясняющих переменных.

Заметим, что нами при построении комитета в качестве зависимой переменной рассматривается только направление цены индекса (рост – 1,

снижение – 0), при этом полностью игнорируется насколько сильно изменилась цена. Так росту цены на 10 EUR и росту цены на 1 EUR будет присвоено одинаковое значение 1.

Рассмотрим возможность учесть «силу» изменения стоимости индекса при применении метода комитетов. Для этого введем новые переменные  $S_i$  и  $S_j$ , значение которых будет определяться по следующей формуле:

$$S_i = \frac{\Delta P_i - \Delta P_{min}}{\Delta P_{max} - \Delta P_{min}} \quad (44)$$

Где  $i$  – порядковый номер наблюдения,  $\Delta P_i$  – модуль изменения стоимости индекса CAC40 за 10 минут в наблюдении  $i$ ,  $\Delta P_{min}$  – минимальный модуль изменения индекса за 10 минут в изучаемом периоде (рассчитывается по обучающей выборке),  $\Delta P_{max}$  – максимальный модуль изменения индекса за 10 минут в изучаемом периоде. Расчет для  $S_j$  аналогичен расчету для  $S_i$ .

Для учета «силы» изменения стоимости индекса внесем изменения в целевую функцию (28), определенную в главе 2:

$$\min L' * \left( \sum_j y_j * S_j + \sum_i y'_i * S_i \right) + \sum_t \sum_j v_j^t * S_j + \sum_t \sum_i v_i'^t * S_i \quad (45)$$

В данной целевой функции «штрафы» за некорректную классификацию наблюдений с большим модулем изменения индекса будут выше, а «штрафы» за некорректную классификацию наблюдений с незначительным изменением индекса будут ниже. В дальнейшем комитет, построенный с учетом изменения «силы» изменения индекса будем называть  $S$  (*strength*) комитетом, а комитет построенный без учета «силы» изменения индекса будем называть NonS комитетом.

На обучающей выборке приведенной выше, был построен S комитет из 7 членов. Результаты приведены в таблице 45.

Таблица 45. Результаты S комитета на обучающей выборке.

Голосование комитета	Общее число наблюдений	Число верных решений	Число неверных решений	Доля правильных решений	Доверительный интервал (95%) точности
-------------------------	------------------------------	----------------------------	------------------------------	-------------------------------	---

«За» рост	383	207	176	54,0%	$\pm 4,2\%$
«Против» роста	617	338	279	54,9%	$\pm 3,3\%$
<b>Итого:</b>	<b>1000</b>	<b>545</b>	<b>455</b>	<b>54,5%</b>	<b><math>\pm 2,6\%</math></b>

Исходя из таблицы 45, точность S комитетного решения, чем точность NonS комитетного решения. Отметим, что сравнивать данные комитетные решения, только по доле верных решений является некорректным поскольку, очевидно, что S комитет, учитывающий «силу» изменения индекса будет иметь более низкие значения по точности, поскольку малые изменения игнорируются.

Для оценки комитетных решений введем показатель отражающий «доходность». Будем считать, что в случае, если комитетом принято верное решение, то возникает доход в размере модуля изменения, вне зависимости от направления изменения, в случае, если комитетом принято неверное решение, то возникают расходы в размере модуля изменения.

В таблице 46 представлено сравнение 2 комитетных решений по доходности на обучающей выборке.

Таблица 46. Сравнение доходности комитетных решений на обучающей выборке.

Комитет	Доля правильных решений	Доходность
NonS комитет	56,9%	411,7
S комитет	54,5%	495,6

Как видно из таблицы 46, S комитет имеет большую доходность, несмотря на более низкую точность.

Рассмотрим качество S комитетного решения на тестовой выборке, аналогичной той, что использовалась ранее.

Таблица 47. Результаты S комитета на тестовой выборке.

Голосование комитета	Общее число наблюдений	Число верных решений	Число неверных решений	Доля правильных решений
«За» рост	4417	2256	205	51,1%
«Против» роста	7022	3599	3423	51,3%
<b>Итого:</b>	<b>11439</b>	<b>5855</b>	<b>3628</b>	<b>51,2%</b>

Как следует из таблицы 47, S комитет имеет большую точность на тестовой выборке по сравнению с NonS комитетом (51,2% против 50,6%).

Сравним данные методы по показателю доходности.

Таблица 48. Сравнение доходности комитетных решений на тестовой выборке.

Комитет	Доля правильных решений	Доходность
NonS комитет	50,6%	292,8
S комитет	51,2%	1442,7

Не смотря на незначительное различие в точности прогноза на тестовой выборке, S комитет имеет доходность в 4,9 раза выше, чем NonS комитет.

В целом S комитет, представляет собой больший интерес для практического применения при торговле индексом, чем NonS комитет.

Для оценки качества комитетного решения сравним данные метода с методами технического анализа (MACD и EMA) и другим методом классификации – Logit методом. Приведем краткое описание данных методов:

- MACD (Moving Average Convergence/Divergence, схождение/расхождение скользящих средних) — технический индикатор для анализа движения цены на рынках акций, ресурсов и валютном рынке. MACD очень популярен среди технических аналитиков благодаря своей простоте и ясности. MACD равен разности двух экспоненциальных скользящих средних, одна из которых быстрее другой, то есть от более быстрой (с меньшим периодом) скользящей средней



отнимают более медленную (с большим периодом). Смысл такой операции заключается в том, что значение MACD будет положительным, если цены будут расти, а если скорость роста цен будет увеличиваться, то MACD также будет увеличиваться. И наоборот – если цены уменьшаются — то и MACD принимает отрицательное значение. Считается, что при увеличении MACD растет спрос, а при уменьшении — падает.

- Индикатор ЕМА (Exponential Moving Average) - это экспоненциальное скользящее среднее, определяется путем добавления к предыдущему значению скользящего среднего определенной доли текущей цены закрытия. При использовании экспоненциальных скользящих средних больший вес имеют последние цены закрытия. Сигналы индикатора ЕМА - в зависимости от направления линии индикатора - вверх или вниз, нужно покупать или продавать. Так же если цена пересекла график индикатора, то это расценивается как смена направления тренда, и соответственно нужно входить в рынок в сторону направления движения цены.

- Logit метод (логистическая регрессия)

Для построения данных методов используется обучающая выборка (1000 наблюдений за период с 17.04.2014 по 19.05.2014), на которой ранее были построены комитетные конструкции.

Оценка качества методов проводилась на тестовой выборке за период с 19.05.2014 по 17.04.2015, содержащей 11 439 наблюдений.

Таблица 49. Оценка методов на тестовой выборке.

Метод	Доля правильных решений	Доходность
NonS комитет	50,6%	292,8
S комитет	51,2%	1442,7
Logit метод	50,7%	1056,8
EMA	52,5%	424,8
MACD	50,1%	931,1

Из таблицы 49 видно, что S комитет, в котором учитывается «сила» изменения индекса имеет, наибольшую доходность по сравнению со всеми остальными методами, тогда как NonS комитет имеет минимальную доходность.

В целом метод комитетов представляет практический интерес для дальнейшего изучения и практического применения при торговле на фондовом рынке, так как доходность данного метода на тестовой выборке за период около 1 года (с 19.05.2014 по 17.04.2015), содержащей 11 439 наблюдений, превышает доходность иных методов, таких как Logit метод, ЕМА, MACD. Заметим, что наибольший практический интерес представляет S комитет, который учитывает «силу» изменения индекса. Так же отметим, что качество комитетного решения имеет тенденцию к снижению, в связи с чем в случае практического применения необходимо перестраивать комитетное правило с некоторой периодичностью. В дальнейшем планируются исследования, направленные на включение в комитетные конструкции не только динамику самого индекса, но и динамику показателей, которые могут оказать влияние на индекс, таких как курс валюты, стоимость нефти и иных фондовых индексов.

## Заключение

Выполненное автором исследование позволило предложить методический инструментарий оценки кредитного риска заемщика физического лица на основе комитетных конструкций.

Актуальность диссертационного исследования обуславливается следующими факторами:

1. Рост рисков необеспеченного потребительского кредитования в РФ на фоне снижения темпов роста, требует от кредитных организаций применения продвинутых методов классификации заемщиков и оценки кредитного риска.

2. В настоящее время Банком России осуществляется внедрение подходов к оценке качества активов и капитала кредитной организации, определенных в рамках «Международной конвергенции измерения капитала и стандартов капитала: новые подходы» (Базель II). Основными нововведениями Базеля II оценке рисков, состоит в допуске внутренних рейтингов к расчету достаточности капитала кредитной организации (IRB approach). Кредитные организации заинтересованы в разработке моделей оценки риска, которые бы удовлетворяли требованиям положений Базеля II.

Чаще всего к построению рейтинговых моделей применяются методы построения линейного дискриминанта, такие как Logit или Probit методы. Однако, методы, основанные на линейном дискриминанте, не позволяют учитывать нелинейные зависимости параметров, характеризующих заемщика, в связи с чем перед кредитными организациями встает проблема поиска более продвинутых методов классификации.

Комитетные конструкции являются одним из методов, позволяющих проводить классификацию с учетом нелинейных зависимостей переменных, замечу, что данный метод не нашел широкого применения к построению рейтинговых моделей.

Классификационные методы используются в том случае, когда требуется оценить кредитные риски по конкретному заемщику, чаще всего такие методы

оценки требуются кредитной организации в случае принятия решения о работе с конкретным клиентом (выдача новой ссуды, реструктуризация, взыскание задолженности).

Основным плюсом метода комитетов по сравнению с иными продвинутыми методами классификации, такими как например нейронные сети, является наличие содержательной и геометрической интерпретации решения, метод комитетов можно представить, как модель кредитного комитета из  $N$  членов.

Одной из основных причин, по которым метод комитетов, ранее не применялся для построения рейтинговых моделей, является значительная вычислительная сложность построения комитета для решения задач больших размерностей. В рамках данной работы для снижения вычислительной сложности задачи построения комитета применяются метод главных компонент, позволивший уменьшить число параметров, характеризующих заемщика, эволюционный алгоритм, определяющий начальное решение задачи линейного частично-целочисленного программирования, и метод последовательной оптимизации комитетного решения.

В рамках диссертационного исследования получены следующие основные результаты.

1. На основе анализа и обобщения литературы автором была сформулирована задача классификации заемщиков методом комитетов (раздел 2.1.). Задача классификации заемщиков методом комитетов заключается в нахождении такого множества членов комитета (линейных полупространств), применение которого позволяет разделить заемщиков на 2 класса (дефолтные и недефолтные) с минимальной ошибкой классификации. Членом комитета называется некоторое полупространство в пространстве признаков  $w^t$ , комитетом называется некоторое множество членов комитета  $Q = (w^1, w^2, \dots, w^q)$ . Процедура голосования (число членов комитета, голосующих за выдачу кредита) обозначается через  $|\{t | x_j \in w^t\}|$ . Комитетом большинства

называется такой комитет, в котором решение о классификации заемщиков принимается простым большинством членов. В случае если  $|\{t|x_j \in w^t\}| > \frac{q}{2}$ , заемщик классифицируется в 1 класс, в случае, если  $|\{t|x_j \in w^t\}| < \frac{q}{2}$ , заемщик классифицируется во 2 класс. Комитетом единогласия называется такой комитет, в котором решение о классификации должно быть принято всеми членами комитета единогласно. В случае если  $|\{t|x_j \in w^t\}| = q$ , заемщик классифицируется в 1 класс, в случае, если  $|\{t|x_j \in w^t\}| < q$ , заемщик классифицируется во 2 класс.

В рамках диссертационного исследования применяется  $p$ -комитет, в котором изначально не устанавливается число членов, которые должны проголосовать за или против, а определяются в процессе построения комитетного решения. При принятии  $p$ -комитетом из  $q$  членов решения по  $j$  заемщику, в случае, если  $|\{t|x_j \in w^t\}| \geq p$ , то принимается решение о классификации в 1 класс, в случае, если  $|\{t|x_j \in w^t\}| < p$ , принимается решение о классификации во 2 класс, где  $p = [1..q]$ . Использование в диссертационном исследовании  $p$ -комитета обусловлено, тем, что комитеты большинства и единогласия можно представить как частные случаи  $p$ -комитета. Так для комитета большинства является  $p$ -комитетом, где  $p = \frac{q}{2}$ , а комитет единогласия является  $p$ -комитетом, где  $p=q$ .

2. Предложена методология построения комитетных конструкций (комитеты большинства, единогласия,  $p$ -комитета и комитет с неравными весами членов) через решение задач линейного частично-целочисленного программирования, которые ранее не применялись для построения комитетных конструкций.

Член комитета можно представить как линейную гиперплоскость (линейный дискриминант), поверх которой находятся заемщики одного класса, а под ней заемщики другого класса. Соответственно на начальном этапе в разделе

2.3 диссертационного исследования рассматривается сведение задачи построения линейного дискриминанта к задаче линейного частично-целочисленного программирования. Комитетом соответственно является некоторый набор членов комитета, осуществляющих процедуру голосования по комитетному правилу. В работе показано построение комитета большинства и единогласия, далее показано построение  $p$ -комитета, который является более общей формой комитетных решений.

Также в работе рассмотрена возможность построения комитета с неравными весами членов через решение задачи частично-целочисленного программирования. Несмотря на большую дискриминационную способность, комитет с неравными весами не показал практической применимости в связи со значительным увеличением сложности решения задачи.

Наиболее важным при решении задач линейного программирования является определение целевой функции задачи. Автором было предложено использование целевой функции направленной на уменьшение не только числа ошибок комитетного решения в целом, но и каждого из членов комитета в частности. Применение данной функции было обосновано эмпирически.

3. В рамках диссертационного исследования обоснована возможность применения комитетных конструкций для построения рейтинговых моделей. Метод комитетов выдает бинарное решение (1 – класс 1, 0 – класс 2). Бинарный ответ комитета не подходит для построения рейтинговой модели, поскольку в данном случае имелось бы только 2 рейтинга заемщиков, что недостаточно для применения IRB-подхода.

В рамках диссертационной работы выдвинута гипотеза, что с возрастанием числа членов комитета, «проголосовавших» за дефолт по кредитному договору, вероятность дефолта по кредитному договору также возрастает. Результаты исследований на данных по заемщикам, предоставленных кредитной организацией подтверждают данную гипотезу.

В таблице 14 раздела 3.2.1 представлена таблица, согласно которой с увеличением числа членов комитета, голосующих за дефолт заемщика, растет вероятность возникновения дефолта по кредитному договору. Дополнительно отмечаем, что возникновение данной зависимости возникает за счет применения целевой функции, в которой минимизируются не только число ошибок комитетного решения, но и число ошибок каждого из членов комитета.

Также отмечаем, что по данным таблицы 14, наблюдается, что в случае, когда 6 членов комитета проголосовало за возникновение дефолта, доля договоров в дефолте была ниже, чем в случае, когда за возникновение дефолта проголосовало 5 членов комитета (4,9% и 7,2% соответственно), при этом в том случае, когда за дефолт заемщика проголосовало 4 члена комитета, доля дефолта составил 4,2%. Для практического применения целесообразно объединить результаты двух последних групп, что производится при построении рейтинговой модели.

4. Для практического применения комитетных конструкций для принятия решений при выдаче кредита необходимо построить рейтинговые модели, которые позволят на основании параметров заемщика оценить уровень риска по ссуде. Так в разделе 3.2 диссертационной работы показано построение рейтинговых моделей на основе комитетных конструкций, которые позволили количественно оценить кредитные риски. На первоначальном этапе был построен комитет из 2 членов, который разбил все множество заемщиков на 2 подмножества: заемщики с низким уровнем риска (доля заемщиков в дефолте 0,96%) и заемщики с высоким уровнем риска (доля заемщиков в дефолте 4,31%). Разбиение на 2 подмножества позволило снизить вычислительную сложность дальнейших расчетов, поскольку для обоих подмножеств можно построить комитеты с меньшим числом членов без потери качества дискриминации. В дальнейшем для каждого из подмножеств были построены  $p$ -комитеты из 7 членов, которые были преобразованы в некоторые рейтинговые модели.

Для заемщиков с низким уровнем риска построена рейтинговая модель из 4 рейтингов, в которой вероятность дефолта варьировалась от 0,51% до 2,03%. При

этом на группу с минимальным уровнем риска приходилось 32,0% от общего числа наблюдений в группе заемщиков с низким уровнем риска, а на группу заемщиков с максимальным уровнем риска приходилось 3,53% от общего числа наблюдений в группе заемщиков с низким уровнем риска.

Для заемщиков с высоким уровнем риска построена рейтинговая модель, состоящая из 5 рейтингов, в которой вероятность дефолта варьировалась от 1,91% до 12,75%. Наибольшую группу (46,59% от общего числа заемщиков с высоким уровнем риска) составили заемщики, за дефолт которых проголосовало 2 члена комитета, доля кредитных договоров в дефолте по данной рейтинговой группе составило 46,59%.

Две рейтинговые модели были объединены в единую авторскую рейтинговую модель, состоящую из 8 групп, при этом заемщики с максимальным риском по группе заемщиков с низким уровнем риска и заемщики с минимальным уровнем риска по группе заемщиков с высоким уровнем риска объединены в одну группу, поскольку доли дефолтов в данных группах сопоставимы.

5. Проведены процедуры валидации рейтинговой модели. Так значение скорректированного индекса Херфиндаля (уровня концентрации) составило 7,2% (при допустимом значении для принятия рейтинговой модели – не более 30%, чем меньше тем лучше), значение мер дискриминационной силы AUROC и AR составило 72,3% и 45,9% соответственно (допустимые значения – более 70% и 40% соответственно, чем больше тем лучше). В целом рейтинговая модель удовлетворяет основным валидационным тестам. Отметим, что полученная рейтинговая модель, превосходит по мерам дискриминационной силы модели AUROC и AR модель, построенную с применением Logit метода.

6. В 3 главе диссертационного исследования проведена оценка экономического эффекта от внедрения рейтинговой модели:

- а. Снижение уровня кредитного риска за счет отказа от кредитования рискованных заемщиков позволит снизить потери кредитной организации от риска.



- б. Дифференциация заемщиков по риску позволяет установить низкие ставки для менее рискованных заемщиков и повышенные ставки для рискованных, что даст конкурентное преимущество на рынке потребительского кредитования.
- с. Снижение требований к капиталу после перехода к оценке капитала на основе IRB подхода позволит увеличить рентабельность капитала и нарастить кредитный портфель без привлечения дополнительного капитала.

7. Разработаны основы методологии применения рейтинговой модели, построенной на основе комитетных конструкций. Очевидно, что с течением времени, качество рейтинговой модели может снизиться, в связи с чем одним из условий применения рейтинговых моделей оценки кредитного риска является наличие методологии, позволяющей корректировать/перестраивать рейтинговую модель с течением времени и новой информацией о дефолтах. Также необходимым условием применения рейтинговой модели является корректировка уровней риска групп, исходя из текущего уровня рисков, зависящего от экономической ситуации.

8. В разделе 3.5 рассмотрена возможность применения инструментария на основе комитетных конструкций к задачам прогнозирования финансовых инструментов. В целом метод комитетов представляет практический интерес для дальнейшего изучения и практического применения при торговле на фондовом рынке, так как доходность данного метода на тестовой выборке за период около 1 года (с 19.05.2014 по 17.04.2015), содержащей 11 439 наблюдений, превышает доходность иных методов, таких как Logit метод (метод классификации), ЕМА, MACD (методы технического анализа).

Практические результаты данного исследования в части применения комитетных конструкций к оценке кредитного риска заемщика физического лица были внедрены в систему оценки риска ОАО «УралТрансБанк» и использованы для формирования стратегии на рынке потребительского кредитования.

Результаты исследования частично использованы при формировании бизнес-плана по расходам на резервы по кредитному портфелю физических лиц в Уральском банке ПАО «Сбербанк».

По мнению автора, одним из перспективных направлений исследования является формирование моделей принятия решений на основе комитетных конструкций, позволяющих улучшить качество работы с действующим кредитным портфелем, в том числе разработка модели сбора просроченной задолженности и модели предложения клиентам дополнительных банковских услуг. Также представляет интерес разработка моделей управления привлеченными средствами кредитной организации, в частности определение групп клиентов, пролонгирующих вклады, и групп клиентов, досрочно закрывающих вклады. Одним из возможных направлений дальнейших исследований является разработка моделей прогнозирования фондовых индексов и курсов валют.

## **Библиографический список**

### **Законы и другие нормативные акты**

1. Международная конвергенция измерения капитала и стандартов капитала: Уточненные рамочные подходы / Базельский комитет по банковскому надзору // Банк международных расчетов. — М.: Банк России, 2004.
2. Международная конвергенция измерения капитала и стандартов капитала: новые подходы / Базельский комитет по банковскому надзору // Банк международных расчетов. — М.: Банк России, 2004.
3. Положение Банка России от 26 марта 2004 г. N 254-П "О порядке формирования кредитными организациями резервов на возможные потери по ссудам, по ссудной и приравненной к ней задолженности".
4. Положение Банка России от 28 декабря 2012 г. № 395-П «О методике определения величины собственных средств (капитала) кредитных организаций ("Базель III")».
5. Инструкция Банка России от 3 декабря 2012 г. N 139-И "Об обязательных нормативах банков".
6. Письмо Банка России от 29 декабря 2012 г. № 192-Т "О Методических рекомендациях по реализации подхода к расчету кредитного риска на основе внутренних рейтингов банков".
7. Письмо Банка России от 23 июня 2004 г №70-Т "О типичных банковских рисках"

### **Монографии**

8. Вольский В.И., Лезина З.М. Голосование в малых группах. Процедуры и методы сравнительного анализа. — М.: Наука. — 1991. — 192 с.
9. Ежов А.А., Шумский С.А. Искусственные нейронные сети в экономике и бизнесе. — М.: — 1998. - 222 с.
10. Еремин И.И. Противоречивые модели оптимального планирования. — М.: Наука. — 1988. — 160 с.

- 11.Еремин И.И. Системы линейных неравенств и линейная оптимизация. – Екатеринбург: УрО РАН. – 2007.
- 12.Еремин И.И., Астафьев Н.Н. Введение в теорию линейного и выпуклого программирования. – М.: Изд-во физ.-мат. лит-ры. – 1975.
- 13.Еремин И.И., Мазуров В.Д., Астафьев Н.Н. Несобственные задачи линейного и выпуклого программирования. М.: Наука. – 1983. – 336 с.
- 14.Лопатников Л. И. Экономико-математический словарь: Словарь современной экономической науки. — 5-е изд., перераб. и доп. — М.: Дело. – 2003.
- 15.Лорьер Ж.-Л. Системы искусственного интеллекта. – М.: Мир. 1991.
- 16.Мазуров В.Д. Метод комитетов в задачах оптимизации и классификации. – М.: Наука. – 1990.
- 17.Просалова В.С. Проблемы оценки кредитоспособности клиентов коммерческих банков. – Владивосток: Издательство ВГУЭС - 2008
- 18.Саймон Х., Нейронные сети: полный курс – М.: Издательский дом Вильямс – 2008
- 19.Шапиро Л., Стокман Дж. Компьютерное зрение. – М.: Бином. Лаборатория знаний. – 2006.

#### **Статьи и тезисы докладов**

- 20.Васильева Е.Е. Ретроспектива подходов к оценке кредитного риска: Базель I, II, III [текст] / Е.Е. Васильева // Проблемы современной экономики. – 2015. – №2 (54).
- 21.Виленкин Н.Я., Шрейдер Ю.А. Принятие экспертных решений на основе мажоритарных структур [текст] / Виленкин Н.Я., Шрейдер Ю.А. // Семиотика и информатика. М.: ВИНТИ. – 1977. – №8. – С. 83–90.
- 22.Власов В.А. Анализ ограничений риска в банковском секторе [текст] / Власов В.А. // Деньги и кредит. 2005. – №2. – с.48-50.
- 23.Волкова О.Б. Исследование модели оценки банковского кредитного риска [текст] / Волкова О.Б. // Вестник ЧГУ. – 2010. – №4.

- 24.Глушкова А.А., Помазанов М.В. Некоторые актуальные проблемы оценки кредитного риска в банковской сфере // Вестник Финансового университета. 2013. №1. С.15-26.
- 25.Завьялов С.О. Оценочные подходы к методам анализа кредитоспособности заемщика [текст] / Завьялов С.О. // Бизнес в законе. – 2011. – №2.
- 26.Зайцева О.А. Базель II. Первый компонент — стандартизированный подход к оценке кредитного риска [текст] / О.А. Зайцева // Документы и комментарии. – 2007. – № 2.
- 27.Карпова Е.С. Экономический и регулятивный капитал в российской банковской практике [текст] / Карпова Е.С. // Universum: экономика и юриспруденция. – 2015. – №5 (16).
- 28.Кахриманова К.Р. Особенности российского банковского надзора и регулирования с точки зрения внедрения Базеля II и Базеля III в российский банковский сектор [текст] / Кахриманова К.Р. // МИР - Модернизация. Инновации. Развитие. – 2014. – №2 (18).
- 29.Кибзун А. И., Нурминский Е. А., Хачай М. Ю. Современные проблемы математического программирования [текст] / Кибзун А. И., Нурминский Е. А., Хачай М. Ю. // Автомат и телемех. – 2012. – № 2. – с. 3-4
- 30.Козлов В.Н. Разпознавание изображений, представляемых конечным множеством точек [текст] / Козлов В.Н. // Фун- дам. прикл. математика. – 2009. Т.15, № 5. – С. 95–110.
- 31.Кривоногов А.И. О некоторых комитетных конструкциях классификации / Кривоногов А.И. // Методы оптимизации и распознавания образов в задачах планирования. Свердловск: УНЦ АН СССР, 1980. С.92-98.
- 32.Кузнецов А.В. Вычислительный эксперимент в экономике и нейронные сети / Обзорение прикладной и промышленной математики. Кисловодск [текст] / Кузнецов А.В. // тезисы доклада Седьмого всероссийского симпозиума по прикладной и промышленной математике 2-8 мая 2006г. – 2006. – Том 13, №3. – с.485.

33. Куштуев А.А. Использование показателей финансовой устойчивости при анализе кредитоспособности клиентов банка [текст] / Куштуев А.А. // Деньги и кредит. - 1998. - №1. - С.30-34.
34. Мазуров В.Д. Модели интерпретации противоречивых данных и метод комитетов [текст] / Мазуров В.Д. // Тр. ИММ УрО РАН. - 1992. - №1. - с.193-203
35. Мазуров В.Д. Об одном методе безусловной оптимизации в задачах выпуклого программирования [текст] / Мазуров В.Д. // Ж. вычисл. матем. и матем. физ. - 1967. - №7(1). - с. 204-208.
36. Мазуров В.Д. Об экспоненциальном методе решения системы выпуклых неравенств [текст] / Мазуров В.Д. // Ж. вычисл. матем. и матем. физ. - 1966. - №6(2). - с. 342-347.
37. Мазуров В.Д. Релаксационный метод в условиях противоречивой системы линейных неравенств (множество предельных точек) [текст] / Мазуров В.Д. // Матем. Заметки. - 1969. - №5(4). - с. 449-456
38. Мазуров В.Д. Коллективное поведение и математическая психология [текст] / Мазуров В.Д. // Экономическое развитие в современном мире. Екатеринбург: Изд-во Урал. гос. ун-та. - 2009. - С. 112-116.
39. Мазуров В.Д. Комитетные решения задач планирования [текст] / Мазуров В.Д. // Методы аппроксимации несобственных задач математического программирования. Свердловск: УНЦ АН СССР. - 1984. - С. 21-25.
40. Мазуров В.Д. Комитеты систем неравенств и задача распознавания образов [текст] / Мазуров В.Д. // Кибернетика. - 1971. - № 3. - С. 140-146.
41. Мазуров В.Д. Распознавание образов как средство автоматического выбора процедуры в вычислительных методах [текст] / Мазуров В.Д. // Вычисл. матем. и матем. физ. - 1970. - №10(6). - с.1520-1525
42. Мазуров В.Д., М. Ю. Хачай, А. И. Рыбин. Комитетные конструкции для решения задач выбора, диагностики и прогнозирования [текст] / Мазуров В.Д., Хачай М.Ю. // Тр. ИММ УрО РАН. - 2002. - №8(1). - с. 66-102

- 43.Мазуров В.Д., Смирнов А.И. Об алгебраическом подходе к восстановлению объектов по их изображениям [текст] / Мазуров В.Д., Смирнов А.И. // Автоматизир. системы обработки изображений (АСОИЗ-86): тез. докл. Всесоюз. конф. (Львов). М.: Наука. – 1986. – С. 154.
- 44.Мазуров В.Д., Смирнов А.И. Противоречивая интерпретация неоднозначных сцен [текст] / Мазуров В.Д., Смирнов А.И. // Методы мат. программирования и их программ. обеспечение: тез. докл. науч.-техн. конф. / Ин-т математики и механики УНЦ АН СССР. Свердловск. – 1984. – С. 72–73.
- 45.Мазуров В.Д., Хачай М.Ю. Бустинг и полиномиальная аппроксимируемость задачи о минимальном аффинном разделяющем комитете [текст] / Мазуров В.Д., Хачай М.Ю. // Тр. ИММ УрО РАН. – 2013. – №19(2). – с.231–236
- 46.Мазуров В.Д., Хачай М.Ю. Комитетные конструкции [текст] / Мазуров В.Д., Хачай М.Ю. // Известия Уральского университета. Сер. Математика–механика. – 1999. – № 2 (14). – С. 77–109.
- 47.Мазуров В.Д., Хачай М.Ю. Комитетные конструкции как обобщение решений противоречивых задач исследования операций [текст] / Мазуров В.Д., Хачай М.Ю. // Дискретн. анализ и исслед. опер., сер. 2. – 2003. – №10(2). – с.56–66.
- 48.Мазуров В.Д., Хачай М.Ю. Комитеты систем линейных неравенств [текст] / Мазуров В.Д., Хачай М.Ю. // Автомат. и телемех. – 2004. – № 2. – с.43–54
- 49.Мазуров В.Д., Хачай М.Ю., Поберий М.И. Задачи комбинаторной оптимизации, связанные с полиэдральной комитетной отделимостью конечных множеств [текст] / Мазуров В.Д., Хачай М.Ю. // Тр. ИММ УрО РАН. – 2008. – № 14(2) (2008). – с.89–102.
- 50.Мазуров Вл.Д., Кривоногов А.И., Казанцев В.С., Сачков Н.О., Белецкий Н.Г. Комитеты в принятии решений [текст] / Мазуров Вл.Д., Кривоногов А.И., Казанцев В.С., Сачков Н.О., Белецкий Н.Г. // Кибернетика – 1984 - №1 – с.90-95.

- 51.Маккалох Дж., Питтс У. Логические исчисления идей, относящихся к нервной деятельности. – Автоматы. М.: ИЛ. – 1956.
- 52.Моисеев А.В. Сравнительный анализ моделей распознавания риска [текст] / А.В.Моисеев, Е.А. Поправко, Н.Г. Федотов. // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2013. – №4 (28). – С. 19-31.
- 53.Монтлевич В.М., Исмаилова А.Н. Эмпирический анализ приближенных алгоритмов целочисленного программирования, основанных на идее жадного выбора [текст] / Монтлевич В.М., Исмаилова А.Н. // Вестник СамГУ — Естественнонаучная серия. – 2014. – № 3(114).
- 54.Мязова Я. С. Сравнительный анализ методик оценки кредитного риска заемщика [текст] / Мязова Я. С. // Известия Самарского научного центра РАН. – 2006. – №4.
55. Никонов. О.И. О нетрадиционных задачах портфельного управления / Никонов. О.И. // Труды института математики и механики УрО РАН. – 2009 - Т.15. №4. С. 204-214.
- 56.Никонов О.И., Медведев М.А. Диверсификация рисков предприятия при взаимодействии с внешними контрагентами / Никонов О.И., Медведев М.А. // Вестник УГТУ-УПИ. Сер. Экономика и управление. 2007. № 1(84).- С.68-72.
- 57.Никонов О.И., Власов В.Е. Оценка эффективности операций корпоративного кредитования с учетом риска / Никонов О.И., Власов В.Е. // Вестник УрФУ. Серия: Экономика и управление. 2013. № 4. С. 127-139.
- 58.Никонов О. И., Чернавин Ф. П. Построение рейтинговых групп заемщиков – физических лиц с применением метода комитетов / О. И. Никонов, Ф.П. Чернавин // «Деньги и Кредит» - 2014. №11 с. 52-55.
59. Никонов О. И., Чернавин Ф. П. Динамика показателя «вероятность дефолта» по кредитам физических лиц / О. И. Никонов, Ф.П. Чернавин // «Деньги и Кредит» - 2015. №2 с. 40-45.



60. Никонов О. И., Чернавин Ф. П. Современные методы классификации. Метод комитетов / О. И. Никонов, Ф.П. Чернавин // Материалы конференции «Социально-экономическое развитие регионов России», 23 мая 2014 года
61. Никонов О. И., Чернавин Ф. П., Власов В. Е. Сравнение метода комитетов и логит-модели в приложениях к решению задач прогнозирования валютного рынка / О. И. Никонов, Ф.П. Чернавин, В. Е. Власов // Материалы X Международной научной конференции по проблемам экономического развития в современном мире «Устойчивое развитие российских регионов: Россия и ВТО», май 2013
62. Никонов О. И., Чернавин Ф. П. Применение метода комитетов к оценке рисков / О. И. Никонов, Ф.П. Чернавин // Материалы конференции XIII международной молодежной конференции "Новые тенденции в экономике и управлении организацией", июнь 2014
63. Никонов О. И. Построение рейтинговых моделей с применением комитетных конструкций / О. И. Никонов, Ф. П. Чернавин, Н. П. Чернавин // Устойчивое развитие российских регионов: экономическая политика в условиях внешних и внутренних шоков : сборник материалов XII международной научно-практической конференции, г. Екатеринбург, 17-18 апреля 2015 г.
64. Никонов О. И. Проблемы классификации: метод комитетов / О. И. Никонов, Ф. П. Чернавин, М. А. Медведева // Устойчивое развитие российских регионов: экономическая политика в условиях внешних и внутренних шоков : сборник материалов XII международной научно-практической конференции, г. Екатеринбург, 17-18 апреля 2015 г.
65. Никонов О.И., Чернавин Ф.П., Чернавин Н.П. Применение метода комитетов к решению задач прогнозирования валютного рынка / О. И. Никонов, Ф. П. Чернавин, Н. П. Чернавин / Вестник УИЭУиП №2 2015.
66. Никонов О.И., Тимофеев Н.А. Поток платежей кредитного портфеля в условиях неполной статистической информации / О.И. Никонов, Н.А.

- Тимофеев // Вестник УрФУ. Сер. Экономика и управление. – 2013. – № 2. – С. 106–111.
67. Попова Е. М., Тюрин Е. И. Базель III и эволюция международного банковского регулирования [текст] / Попова Е. М., Тюрин Е. И. // Финансы и бизнес. – 2012. – № 2. – С. 47-48
68. Порошина А.М. Развитие моделей кредитного риска на рынке ипотечного кредитования [текст] / Порошина А.М. // УЭКС. – 2012. – №48 (12). – С.81.
69. Рустамов С.А. Херштаттский риск (рус.) [текст] / Рустамов С.А. // Валютный спекулянт: журнал. – 2003. – № 5.
70. Селянин В. Е., Андрейчиков А. В. Концепция методики использования нечётких нейронных сетей для оценки кредитного риска на рынке потребительского кредитования [текст] / Селянин В. Е., Андрейчиков А. В. // Известия ВолгГТУ. – 2006. – №6. – С.228-234.
71. Симановский А.Ю. Банковское регулирование: реэволюция (Часть I) [текст] / Симановский А.Ю. // Деньги и кредит. – 2014. – №3.
72. Симановский А.Ю. Банковское регулирование: реэволюция (Часть II) [текст] / Симановский А.Ю. // Деньги и кредит. – 2014. – №9.
73. Симановский А.Ю. Достаточность капитала: еще раз к концепции [текст] / Симановский А.Ю. // Деньги и кредит. – 2014. – №4.
74. Симановский А.Ю. Кризис и реформа регулирования: отдельные аспекты [текст] / Симановский А.Ю. // Деньги и кредит. – 2014. – №12.
75. Симановский А.Ю. О регулярных требованиях к устойчивости банков (развернутые тезисы) [текст] / Симановский А.Ю. // Деньги и кредит. – 2009. – №9.
76. Симановский А.Ю. Регулятивные требования к капиталу: возможны ли альтернативы? [текст] / Симановский А.Ю. // Деньги и кредит. – 2008. – №7.
77. Смирнов А. И. Интерпретация противоречивых изображений на основе систем линейных неравенств [текст] / Смирнов А. И. // Тр. ИММ УрО РАН. – 2012. – №18(3). – с. 144–154

- 78.Стрельников Е.В. Проблемы применения метода RAROC при расчете потребности в капитале / Стрельников Е.В. // Современные проблемы науки и образования. – 2012. – № 4.
- 79.Тотьмянина К.М. Обзор моделей вероятности дефолта, управление финансовыми рисками [текст] / Тотьмянина К.М. // 2011. – № 01(25).
- 80.Травкина Е.В. Внедрение базельских соглашений в российскую банковскую систему [текст] / Травкина Е.В. // Известия ОГАУ. – 2012. – №35-1. – С.179-182.
- 81.Урсуленко А. В. Базель II: Особенности моделирования кредитного риска [текст] / Урсуленко А. В. // Всероссийский журнал научных публикаций. – 2010. – №1 (1)
- 82.Хачай М. Ю. Вопросы вычислительной сложности процедур обучения распознаванию в классе комитетных кусочно-линейных решающих правил [текст] / Хачай М. Ю. // Автомат. и телемех. – 2010. – № 3. – с.178–189
- 83.Хачай М.Ю. О вычислительной и аппроксимационной сложности задачи о минимальном аффинном комитете [текст] / Хачай М.Ю. // Таврич. вест. информ. и мат. – 2006. – № 1. – С. 34–43.
- 84.Хачай М.Ю. О вычислительной сложности задачи о минимальном комитете и смежных задач [текст] / Хачай М.Ю. // ДАН. 2006. – Т. 406. – № 6. – С. 742–745.
- 85.Хачай М.Ю. Об одной игре с природой, связанной с принятием решений большинством голосов [текст] / Хачай М.Ю. // Журнал вычислит. матем. и матем. физики. – 2002. – Т. 42. – № 10. – С. 1609–1616.
- 86.Хачай М.Ю. Об одном соотношении, связанном с процедурой принятия решений большинством голосов [текст] / Хачай М.Ю. // ДАН. – 2001. – Т. 381, № 6. – С. 748–752.
- 87.Хачай М.Ю., Поберий М.И. Вычислительная сложность задач комитетной полиэдральной отделимости в пространствах фиксированной размерности [текст] / Хачай М.Ю., Поберий М.И. // Таврич. вест. информ. и мат. – 2008. – № 2. С. – 218–227.

88. Чамокова Ф. А. Базель II: методика оценки и управления кредитным риском [текст] / Чамокова Ф. А. // Новые технологии. – 2007. – №4.
89. Чернавин Ф.П. Применение комитетных конструкций для принятия решений по потребительским кредитам / Ф.П. Чернавин // «Экономика и предпринимательство» - 2015 №12-4(65-4) с. 143-149
90. Чернавин Ф. П. Применение нейронных сетей к задачам оценки вероятности дефолта по потребительским кредитам/ Ф.П. Чернавин // Журнал научных публикаций аспирантов и докторантов, июль 2013 года
91. Юткин С.С., Туйсузов А.В. Оценка кредитного риска на базе факторного анализа методом IRB Approach [текст] / Юткин С.С., Туйсузов А.В. // МНИЖ. – 2015. – №8-1 (39).

#### **Интернет-источники**

92. Altman E. Default Recovery Rates and LGD in Credit Risk Modeling and Practice: An Updated Review of the Literature and Empirical Evidence. 2006 // [Electronic source]. – Access mode: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.7082&rep=rep1&type=pdf>
93. Eric Rosenberg, Alan Gleit. Quantitative Methods in Credit Management: A Survey // [Electronic source]. – Access mode: <http://pubsonline.informs.org/doi/abs/10.1287/opre.42.4.589>
94. Maria Stepanova, Lyn Thomas. Survival Analysis Methods for Personal Loan Data // [Electronic source]. – Access mode: <http://pubsonline.informs.org/doi/abs/10.1287/opre.50.2.277.426>.
95. Rudiger Frey, Alexander J. McNeil. Dependent Defaults in Models of Portfolio Credit Risk. 2003 // [Electronic source]. – Access mode: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.7946&rep=rep1&type=pdf>
96. Sarah Bridges, Richard Disney. Modelling Consumer Credit and Default: The Research Agenda // [Electronic source]. – Access mode:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.194.2879&rep=rep1&type=pdf>

- 97.Симановский А.Ю. Перспективы банковского регулирования: отдельные аспекты. [Электронный ресурс]. – Режим доступа: [http://www.cbr.ru/publ/MoneyAndCredit/Simanovski\\_0709.pdf](http://www.cbr.ru/publ/MoneyAndCredit/Simanovski_0709.pdf)
- 98.Померанцев А Метод Главных Компонент (PCA) [Электронный ресурс]. - Режим доступа: <http://www.chemometrics.ru/materials/textbooks/pca.htm>
- 99.Валидация. Комитет АРБ по стандартам Базель II и управлению рисками [Электронный ресурс]. - Режим доступа: <http://arb.ru/upload/iblock/6da/%D0%92%D0%B0%D0%BB%D0%B8%D0%B4%D0%B0%D1%86%D0%B8%D1%8F.docx>
- 100.Аналитический документ о степени соответствия внутрибанковских подходов к управлению кредитным риском банков – участников проекта «Банковское регулирование и надзор (Базель II)" Программы сотрудничества Евросистемы с Банком России минимальным требованиям IRB-подхода Базеля II [Электронный ресурс]. - Режим доступа: <http://www.cbr.ru/today/ms/bn/GAP.pdf>

### **Иностранная литература**

- 101.Ablow, CM. and Kaylor, D.J., Inconsistent Homogenous Linear Inequalities // Bulletin of the American Mathematical Society, 1965, vol. 71, no 5, p. 724.
- 102.Banzhaf, W., Nordin, P., Keller, R.E., Francone, F.D. (1997), Genetic Programming: An Introduction: On the Automatic Evolution of Computer Programs and Its Applications, Morgan Kaufmann
- 103.Hand D. J., Henley W. E.. Statistical Classification Methods in Consumer Credit Scoring: A Review. Journal of the Royal Statistical Society. Series A (Statistics in Society) Vol. 160, No. 3 (1997), pp. 523-541
- 104.Khashman A. Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. Expert Systems with Applications Volume 37, Issue 9, September 2010, Pp. 6233–6239
- 105.Allen L., Saunders A. A Survey of Cyclical Effects in Credit Risk Measurement Models. NYU Working Paper No. FIN-02-018 May 2002

106. Altman E., Resti A., Sironi A. Default Recovery Rates in Credit Risk Modelling: A Review of the Literature and Empirical Evidence. *Economic Notes* Volume 33, Issue 2, pp. 183–208, July 2004.
107. Altman E., Saunders A. Credit Risk Measurement: Developments over the Last 20 Years (Digest Summary). *Journal of Banking & Finance*, 1997, Vol. 21, No. 11 and 12
108. Amir E. Khandani, Adlar J. Kim, Andrew W. Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance* Volume 34, Issue 11, November 2010, Pp. 2767–2787
109. Angelini E., Tollo G. A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, Volume 48, Issue 4, November 2008, pp. 733–755
110. Anthony D. Miyazaki, Ana Fernandez. Consumer Perceptions of Privacy and Security Risks for Online Shopping. *Journal of Consumer Affairs* Volume 35, Issue 1, pp. 27–44, 2001
111. Arminger G., Universitat B., Enache D., Universitat B., Bonne T. Analyzing Credit Risk Data: A Comparison of Logistic Discrimination, Classification Tree Analysis, and Feedforward Networks. *Computational statistics*, Vol 12 No. 2, March 26, 1997
112. Atiya, A.F. Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, Volume 12, Issue 4, 2002
113. Kamleitner B., Kirchler E.. Consumer credit use: a process model and literature review. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* Volume 57, Issue 4, 2007, pp. 267–283
114. Barry Z. Cynamon, Steven M. Fazzari. Household Debt in the Consumer Age: Source of Growth--Risk of Collapse. *Capitalism and Society*, Volume 3, Issue 2, October 2008
115. Bessis, J. Risk management in banking / J. Bessis. - 2nd ed. – Chichester : John Wiley & Sons Ltd., 1998.

116. Bhekisipho Twala. Multiple classifier application to credit risk assessment. *Expert Systems with Applications* Volume 37, Issue 4, April 2010, Pp. 3326–3336.
117. Careya M., Hrycayb M. Parameterizing credit risk models with rating data. *Journal of Banking & Finance* Volume 25, Issue 1, January 2001, Pp. 197–270
118. Carling K., Jacobson T., Lindé J., Roszbach K. Capital Charges under Basel II: Corporate Credit Risk Modelling and the Macro Economy. Sveriges Riksbank Working Paper Series 142, 2002.
119. Chorng-Shyong Onga, Jih-Jeng Huanga, Gwo-Hshiung Tzengb. Building credit scoring models using genetic programming. *Expert Systems with Applications* Volume 29, Issue 1, July 2005, Pp. 41–47
120. Christian Bluhm, Ludger Overbeck, Christoph Wagner. *Introduction to Credit Risk Modeling* (second edition). 2010
121. Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer Science+Business Media, LLC, 2006.
122. Churlzu Lim · Hanif D. Sherali · Stan Uryasev, *Portfolio optimization by minimizing conditional value-at-risk via nondifferentiable optimization*, Springer Science+Business Media, LLC 2008,
123. Crooka J.N., Edelmanb D.B., Thomasc L.C. Recent developments in consumer credit risk assessment, *European Journal of Operational Research* Volume 183, Issue 3, 16 December 2007, Pp. 1447–1465.
124. Crouhy M., Galaib D., Mark R. A comparative analysis of current credit risk models. *Journal of Banking & Finance* Volume 24, Issues 1–2, January 2000, Pp. 59–117
125. Hand D J. Good practice in retail credit scorecard assessment. *Journal of the Operational Research Society* 56, 2005
126. Dan J. Kima, Donald L. Ferrinb, H. Raghav Raoc. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems* Volume 44, Issue 2, January 2008, pp. 544–564

127. Daniel Rösch, Harald Scheule. The Journal of Risk Finance, Vol. 5 Iss 2 pp. 16 - 32, 2004
128. Darrell Duffie. Innovations in Credit Risk Transfer: Implications for Financial Stability. BIS Working Paper No. 255, July 1, 2008.
129. Darryl E. Getter. Consumer Credit Risk and Pricing. Journal of Consumer Affairs Volume 40, Issue 1, pp. 41–63, Summer 2006
130. David J. Hand. Modelling consumer credit risk. IMA J Management Math (2001) 12 (2): 139-155.
131. David West, Neural network credit scoring models, Computers & Operations Research Volume 27, Issues 11–12, September 2000, pp. 1131–1152
132. Dean Fantazzini, Silvia Figini. Random Survival Forests Models for SME Credit Risk Measurement. Methodology and Computing in Applied Probability March 2009, Volume 11, Issue 1, pp. 29-45
133. Donncha M.. Lending by numbers: credit scoring and the constitution of risk within American consumer credit. Economy and Society Volume 36, Issue 1, 2007
134. Euromoney Books, SBC Warburg Dillon Read, Goldman Sachs. The Practice of Risk Management, Hardcover – February, 1998
135. Fabio Wendling Muniz de Andradea, Abraham Laredo Sicsúb. A Credit Risk Model for Consumer Loan Portfolios. Latin American Business Review Volume 8, Issue 3, 2008
136. Fabio Wendling Muniz de Andradea, Lyn Thomasb. Structural models in consumer credit. European Journal of Operational Research Volume 183, Issue 3, 16 December 2007, Pp. 1569–1581
137. G. Verstraeten, D. Van den Poel. The impact of sample bias on consumer credit scoring performance and profitability. Journal of the Operational Research Society No 56, pp. 981–992, 2005
138. Gang Wanga, Jinxing Haob, Jian Mab, Hongbing Jiangb. A comparative assessment of ensemble learning for credit scoring. Expert Systems with Applications Volume 38, Issue 1, January 2011, Pp. 223–230.



139. Goodhart C., Segoviano A. Basel and procyclicality: a comparison of the standardised and IRB approaches to an improved credit risk method. Discussion paper, 524. Financial Markets Group, London School of Economics and Political Science, London, UK, 2004
140. H. Zhu, P. A. Beling, G. A. Overstreet. A Study in the Combination of Two Consumer Credit Scores. The Journal of the Operational Research Society Vol. 52, No. 9, Special Issue: Credit Scoring and Data Mining (Sep., 2001), pp. 974-980
141. Hamerle A., Liebig T., Rösch D. Credit Risk Factor Modeling and the Basel II IRB Approach. Discussion Paper Series 2: Banking and Financial Supervision, N 02/2003.
142. Hussein Abdou, John Pointon, Ahmed El-Masry. Neural nets versus conventional techniques in credit scoring in Egyptian banking. Expert Systems with Applications Volume 35, Issue 3, October 2008, Pp. 1275–1292
143. Hyafil Laurent, Rivest RL. Constructing Optimal Binary Decision Trees is NP-complete. Information Processing Letters. 5 (1) (1976).
144. J. Galindo, P. Tamayo. Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications. Computational Economics April 2000, Volume 15, Issue 1, pp 107-143
145. Jacobson T., Lindé J., Roszbach K. Credit Risk Versus Capital Requirements under Basel II: Are SME Loans and Retail Credit Really Different? Journal of Financial Services Research October 2005, Volume 28, Issue 1, pp 43-75.
146. John Muellbauer. Housing, Credit and Consumer Expenditure. CEPR Discussion Paper No. 6782, 2008.
147. Jonathan N. Crooka, David B. Edelmanb, Lyn C. Thomasc. Recent developments in consumer credit risk assessment. European Journal of Operational Research Volume 183, Issue 3, 16 December 2007, Pp. 1447–1465
148. Kin Keung Lai, Lean Yu, Ligang Zhou, Shouyang Wang. Credit Risk Evaluation with Least Square Support Vector Machine. Lecture Notes in Computer Science, Vol. 4062, pp. 490-495, 2006

- 149.Kin Keung Lai, Lean Yu, Shouyang Wang, Ligang Zhou. Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model. 16th International Conference, Athens, Greece, September 10-14, 2006.
- 150.Kin Keung Lai, Lean Yu, Shouyang Wang, Ligang Zhou. Neural Network Metalearning for Credit Scoring.International Conference on Intelligent Computing, ICIC 2006, Kunming, China, August 16-19, 2006. Proceedings, Part I, 2006
- 151.L. C. Thomas, R. W. Oliver, D. J. Hand. A survey of the issues in consumer credit modelling research. Journal of the Operational Research Society (2005) 56, 1006–1015.
- 152.Lean Yua, Shouyang Wanga, Kin Keung Laib. An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring. European Journal of Operational Research Volume 195, Issue 3, 16 June 2009, Pp. 942–959
- 153.Lean Yua, Shouyang Wanga, Kin Keung Laib. Credit risk assessment with a multistage neural network ensemble learning approach. Expert Systems with Applications Volume 34, Issue 2, February 2008, Pp. 1434–1444
- 154.Lean Yua, Wuyi Yueb, Shouyang Wanga, K.K. Laic. Support vector machine based multiagent ensemble learning for credit risk evaluation. Expert Systems with Applications Volume 37, Issue 2, March 2010, pp. 1351–1360
- 155.Lean Yua, Xiao Yaoa, Shouyang Wanga, K.K. Laic. Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. Expert Systems with Applications, Volume 38, Issue 12, 2011, pp. 15392–15399
- 156.Leung K., Cheong F., Cheong C. Consumer credit scoring using an artificial immune system algorithm. Evolutionary Computation, 2007
- 157.Li-Chiu Chi, Tseng-Chung Tang. Bankruptcy Prediction: Application of Logit Analysis in Export Credit Risks. Australian Journal of Management June 2006 vol. 31 no. 1 17-27.

- 158.Linda Allena, Gayle DeLonga, Anthony Saundersb. Issues in the credit risk modeling of retail markets. *Journal of Banking & Finance* Volume 28, Issue 4, April 2004, Pp. 727–752
- 159.M. Malik, L. C. Thomac. Modelling credit risk of portfolio of consumer loans.
- 160.Maja Šušteršiča, Dušan Mramorb, Jure Zupanc. Consumer credit scoring models with limited data. *Expert Systems with Applications* Volume 36, Issue 3, Part 1, April 2009, Pp. 4736–4744.
- 161.Mark Careya, Mark Hrycayb. Parameterizing credit risk models with rating data. *Journal of Banking & Finance* Volume 25, Issue 1, January 2001, Pp. 197–270
- 162.Mark J. Furletti. An Overview and History of Credit Reporting. FRB of Philadelphia Payment Cards Center Discussion Paper No. 02-07, 2002
- 163.Michael B. Gordy. A comparative anatomy of credit risk models. *Journal of Banking & Finance* Volume 24, Issues 1–2, January 2000, Pp. 119–149
- 164.Michael B. Gordy. A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation* Volume 12, Issue 3, July 2003, Pp. 199–232
- 165.Michael K. Lima, So Young Sohnb. Cluster-based dynamic scoring model. *Expert Systems with Applications* Volume 32, Issue 2, February 2007, Pp. 427–431
- 166.Michel Dietsch, Joël Petey. The credit risk in SME loans portfolios: Modeling issues, pricing, and capital requirements. *Journal of Banking & Finance* Volume 26, Issues 2–3, March 2002, Pp. 303–322
- 167.Mirta Bensic, Natasa Sarlija, Marijana Zekic-Susac. Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management* Volume 13, Issue 3, pp. 133–150, July/September 2005
- 168.Murthy S. Automatic construction of decision trees from data: A multidisciplinary survey. *Data Mining and Knowledge Discovery*, 1998.

169. Pearson K., On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, (1901) 2, 559—572;
170. Paul S. Calem, Loretta J. Mester. Consumer Behavior and the Stickiness of Credit-Card Interest Rates. *The American Economic Review* Vol. 85, No. 5 (Dec., 1995), pp. 1327-1336
171. Peter Burns, Anne Stanley. Managing Consumer Credit Risk. Federal Reserve Bank of Philadelphia Payment Cards Center Discussion Paper No. 01-03, 2001
172. Pierre Collin-Dufresne, Robert S. Goldstein, Jean Helwege. Is Credit Event Risk Priced? Modeling Contagion via the Updating of Beliefs. NBER Working Paper No. 15733, February 2010
173. Quinlan, J. R., Induction of Decision Trees. *Machine Learning* - 1986 - №1: 81-106, Kluwer Academic Publishers
174. R. Fildes, K. Nikolopoulos, S. F. Crone, A. A. Syntetos. Forecasting and operational research: a review. *Journal of the Operational Research Society* No 59, 1150–1172, 2008
175. Rashmi Malhotraa, D.K Malhotrab. Evaluating consumer loans using neural networks. *Omega* Volume 31, Issue 2, April 2003, Pp. 83–96.
176. Richard J. Herring. Credit Risk and Financial Instability. Ratings, Rating Agencies and the Global Financial System, pp 345-367, 2002
177. Robert A. Jarrow, David Lando, Stuart M. Turnbull. A Markov Model for the Term Structure of Credit Risk Spreads. *Oxford Journals, Social Sciences, Review of Financial Studies*, Volume 10, Issue 2, Pp. 481-523. 1997
178. Robert B Avery, Paul S Calem, Glenn B Canner. Consumer credit scoring: Do situational circumstances matter?. *Journal of Banking & Finance* Volume 28, Issue 4, April 2004, Pp. 835–856
179. Satyajit Chatterjee, Philip Dean Corbae, Makoto Nakajima, José-Victor Rios-Rull. A Quantitative Theory of Unsecured Consumer Credit with Risk of Default. FRB Philadelphia Working Paper No. 07-16. 2007

180. Shu Ling Lin. A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications* Volume 36, Issue 4, May 2009, Pp. 8333–8341
181. Sjur Westgaard, Nico van der Wijk. Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research* Volume 135, Issue 2, 1 December 2001, Pp. 338–349
182. Sylvia Lane. Submarginal Credit Risk Classification. *Journal of Financial and Quantitative Analysis*, Volume 7, Issue 01, January 1972, pp 1379-1385
183. Tian-Shyug Lee, Chih-Chou Chiub, Chi-Jie Luc, I-Fei Chend. Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications* Volume 23, Issue 3, 1 October 2002, Pp. 245–254
184. Tian-Shyug Lee, Chih-Chou Chiub, Chi-Jie Luc, I-Fei Chend. Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications* Volume 23, Issue 3, 1 October 2002, pp. 245–254
185. Tony Bellotti, Jonathan Crook. Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications* \Volume 36, Issue 2, Part 2, March 2009, Pp. 3302–3308
186. W. E. Henley, D. J. Hand. A k-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk. *Journal of the Royal Statistical Society. Series D (The Statistician)* Vol. 45, No. 1 (1996), pp. 77-95
187. Wenbing Xiao, Qian Zhao, Qi Fei. A comparative study of data mining methods in consumer loans credit scoring management. *Journal of Systems Science and Systems Engineering*, December 2006, Volume 15, Issue 4, pp 419-435
188. William F Treacy, Mark Carey. Credit risk rating systems at large US banks. *Journal of Banking & Finance* Volume 24, Issues 1–2, January 2000, Pp. 167–201
189. Yongqiao Wang, Shouyang Wang, Lai K.K. A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems*, Volume 13, Issue 6, Dec. 2005.

190.Yoon Seong Kim, So Young Sohn. Managing loan customers using misclassification patterns of credit scoring model. Expert Systems with Applications, Volume 26, Issue 4, May 2004, Pp